

TAMPEREEN YLIOPISTON TUTKIJOIDEN NÄKEMYKSIÄ TUTKIMUSAINEISTOJEN HALLINNASTA

Hanna Isola

Tampereen yliopisto
Viestintätieteiden laitos
Informaatiotutkimus ja interak-
tiivinen media
Pro gradu -tutkielma
Toukokuu 2018

Pro gradu -tutkielmassani tutkin tutkijoiden näkemyksiä ja kokemuksia tutkimusaineistojen aineistohallinnasta. Tutkimukseni tuo esiin tutkijoiden tapoja toimia tutkimusaineistojen kanssa sekä tutkimusaineistojen aineistohallintaan liittyviä ongelmia.

Tutkimusaineistojen hallinnan kehittäminen on ajankohtainen aihe Suomessa, sillä tutkimusaineistojen, tietokantojen ja tilastojen saatavuutta on käsitelty monissa erilaisissa työryhmissä ja hankkeissa. Aineistohallinnan kehittäminen on hyödyksi tutkimuksen edistämiseksi ja sillä on myös taloudellisia ja poliittisia etuja. Aineistojen siirtyminen digitaalseksi on muuttanut tutkimuksen luonnetta. Nykyään tutkimukseen kerättyä materiaalia ei käytetä pelkästään todistamaan yhden tutkimuksen hypoteeseja, vaan aineistoja käytetään myös uusien hypoteesien ja tutkimusten luomiseen.

Tutkimukseni on toteutettu haastattelemalla Tampereen yliopiston tutkijoita. Menetelmänä on käytetty teemahaastattelua. Haastattelin sekä kvalitatiivista että kvantitatiivista tutkimusta tekeviä tutkijoita.

Tutkimuksessa selvisi, että kaiken kaikkiaan aineistohallinta on monimuotoinen aihealue, johon liittyy useita osa-alueita ja näkökulmia. Tutkijoille oli osittain vaikea tunnistaa, mitä kaikkea aineistohallintaan kuuluu, mutta toisaalta aineistohallinta oli heille joiltain osin kuitenkin tuttua. Osalle tutkijoista oli kokemusta aineistohallintasuunnitelman laatimisesta. Koska kaikille ei ollut aivan selvää, mitä aineistohallintasuunnitelmaan tulisi sisällyttää, laadittiin suunnitelmat melko pintapuolisesti.

Aineistohallinnasta puhuttaessa tutkijoille tuli useimmiten mieleen tutkimusaineistojen arkistointi ja jatkokäyttö. Ongelmina nousivat esiin esimerkiksi yksityisyydensuojan turvaaminen etenkin laadullisten aineistojen arkistoinnin ja jatkokäytön osalta. Tutkimuksessa esiin nousseet aineistohallintaan liittyvät ongelmat olivat pitkälti saman tyyppisiä kuin aiemmin tehdyssä tutkimuksessa.

Avainsanat: aineistohallinta, arkisto, asiakirjahallinta, digitaalinen kuratointi, tutkimusaineistot

SISÄLLYSLUETTELO

1	JOHDANTO	1
2	KIRJALLISUUSKATSAUS	2
2.1	Tutkimusaineistot ja aineistohallinta	2
2.1.1	Tutkimusaineistojen elinkaari	6
2.1.2	Aineistohallinnan haasteita.....	10
2.2	Aiempi tutkimus	11
2.2.1	Aineistohallinnan käytäntöjä Britanniassa	12
2.2.2	Tilanne Suomessa	14
2.2.3	Aineistohallinnan suunnitteleminen	15
3	TUTKIMUSMENETELMÄT	19
3.1	Tutkimuskysymykset.....	19
3.2	Tutkimusaineisto ja –menetelmät.....	19
3.2.1	Laadullinen tutkimus ja teemahaastattelu	20
3.2.2	Aineiston kerääminen	22
3.2.3	Aineiston analyysi	23
4	TUTKIMUSTULOKSET	24
4.1	Aineistohallinnan käytännöistä	24
4.1.1	Suunnittelusta	24
4.1.2	Kerätyn aineiston käsittely	28
4.1.3	Arkistointi ja jatkokäyttö	31
4.2	Ongelmista.....	34
4.2.1	Tutkimuksen aikana.....	34
4.2.2	Tekniset ongelmat	36
4.2.3	Arkistoinnin ja jatkokäytön ongelmista.....	37
5	POHDINTAA JA LOPPUPÄÄTELMÄT	43
	LÄHTEET	47

1 JOHDANTO

Pro gradu –tutkielmani aiheena on Tampereen yliopiston tutkijoiden näkemykset tutkimusaineistojen hallinnasta. Tutkimuksen tarkoituksena on selvittää tutkijoiden näkemyksiä ja kokemuksia tutkimusaineistojen hallintaan liittyvistä tavoista ja haasteista.

Tutkimus tuottaa paljon aineistoa ja nykyään pääsääntöisesti lähes kaikki tutkimusaineisto on digitaalisessa muodossa (Ray 2014). Digitaaliset aineistot mahdollistavat tutkimustiedon helpon jakamisen ja uudelleenkäytön, mutta digitaalisuus aiheuttaa myös omat haasteensa. Jotta tutkimustieto olisi saatavilla, tunnistettavissa ja käytettävissä, pitää aineistohallintaa suunnitella jo tutkimuksen alussa ja kiinnittää huomiota myös tutkimusprosessin aikana. (Pryor 2012).

Tutkimusaineistojen hallinta on yhteiskunnallisesti tärkeää. Aineistohallinnan kehittäminen on hyödyksi tutkimuksen edistämiseksi ja sillä on myös taloudellisia ja poliittisia etuja. Aineistojen siirtyminen digitaalseksi on muuttanut tutkimuksen luonnetta. Nykyään tutkimukseen kerättyä materiaalia ei käytetä pelkästään todistamaan yhden tutkimuksen hypoteeseja, vaan aineistoja käytetään myös uusien hypoteesien ja tutkimusten luomiseen. Uudet keksinnöt ja innovaatiot perustuvat usein vanhoihin tutkimustuloksiin, jolloin aineistojen helppo saatavuus on taloudellisesti hyödyllistä. Poliitikassa tutkimusaineistojen parempi saatavuus on puolestaan edistänyt päätöksentekoa ja viranomaisten keskinäistä tiedonvaihtoa. (CSC– Tieteen tietotekniikan keskus Oy 2010).

Tutkimusaineistojen hallinnan kehittäminen on ajankohtainen aihe Suomessa, sillä tutkimusaineistojen, tietokantojen ja tilastojen saatavuutta on käsitelty monissa erilaisissa työryhmissä ja hankkeissa. Esimerkiksi eri ministeriöt ovat tehneet selvityksiä ja raportteja digitaalisen tiedon avoimuudesta, tutkimustiedon jatkokäytön tehostamisesta ja siitä, kuinka tutkimus-, tilasto- ja rekisteritietoa voitaisiin käyttää paremmin digitaalisten palveluiden kehittämisessä. (CSC– Tieteen tietotekniikan keskus Oy 2010).

Pro gradu -tutkielmani tavoitteena on selvittää, minkälaisia tapoja tutkijoilla on tehdä aineistohallintaa ja minkälaisia ongelmia tutkijat ovat kohdanneet aineistohallintaan liittyen. Tutkimus on laadullinen tutkimus ja menetelmänä on käytetty teemahaastattelua.

2 KIRJALLISUUSKATSAUS

Tässä luvussa määrittelen tutkielmani kannalta keskeisiä käsitteitä sekä aineistohallinnasta tehtyä aikaisempaa tutkimusta. Alaluvussa 2.1 määrittelen tutkimusaineistojen sekä aineistohallinnan merkityksiä. Kerron tutkimusaineistojen elinkaaresta sekä aineistohallinnan peruselementeistä. Alaluvussa 2.2 kerron aineistohallinnasta tehtyä aikaisempaa tutkimusta sekä kirjallisuudessa ja lehdistä esiin tullutta keskustelua aiheesta.

2.1 Tutkimusaineistot ja aineistohallinta

Tutkimuksessa kerääntyy pääasiallisesti kahdenlaista tietoa: tutkimusta varten kerätty aineisto ja itse julkaistu tutkimus. Yleisesti näkyvin osa tutkimuksessa syntyvää tietoa on julkaistu tutkimus, jossa kerättyä aineistoa on analysoitu tutkimuskysymyksiin vastaamiseksi. Tutkimustiedon saatavuus on ennen ollut säädellympää. (Pryor 2012). Joissain tapauksissa on esimerkiksi ajateltu, että tutkimusaineisto on tutkimusta rahoittavan tahon omaisuutta (Parry ja Mauthner 2004). Nykyisenä tavoitteena on kuitenkin tutkimustuloksien mahdollisimman avoin saatavuus, jolloin joko organisaatiot tai itse tutkijat laittavat tutkimukset kaikkien saatavilla ja hyödynnettäväksi. (Pryor 2012).

Aikaisemmin tutkimusta varten kerätty aineisto on ollut osittain vain tutkimuksen sivutuotteena syntynyttä aineistoa, jotka tutkija yksin käyttää tutkimuksensa tekoon. Nykyään on kuitenkin havaittu, että tutkimukseen kerätty aineisto on julkaistun tutkimuksen ohella myös tärkeää säilyttää ja mahdollistaa sen uudelleenkäyttö. Kerätyn aineiston uudelleenkäytöllä voidaan puolueettomasti todistaa tutkimuksen väitteet oikeiksi, mutta mikä tärkeintä, käyttää kerättyä aineistoa uuden tiedon luomiseen uusissa tutkimuksissa. (Higgins 2012, 18).

Tutkimusaineistoa on hyvin erilaisissa muodoissa ja niiden käyttö ja säilytysarvot vaihtelevat. Higgins (2012) määrittelee, että tutkimuksessa syntyvä aineisto on binäärisessä muodossa olevaa tietoa ja sitä luodaan, säilötään ja käsitellään tietotekniikan avulla. Tutkimusprosessissa syntyvät digitaalinen aineisto voi yksinkertaisimmillaan olla yksittäisiä teksti-, kuva- tai äänitiedostoja. Toisaalta aineiston muoto voi olla myös yhdistelmä erilaisia digitaalisia objekteja, kuten esimerkiksi nettisivu. (Higgins 2012).

Vuonna 2007 tehdyssä OECD:n aineistosuosituksessa korostetaan tutkimustiedon avoimuutta, saatavuutta ja pitkäaikaista säilyvyyttä. Vaikka suositus koskee julkisrahoitteista tutkimusta, sitä voi soveltaa kaikkeen tutkimukseen. OECD:n suosituksessa vedotaan tutkimuksen rahoittajiin ja tutkimusinstituutioihin tutkimusaineistojen pitkäaikaisessa säilytyksessä. Suosituksen mukaan rahoittajien tulisi huolehtia uusien tutkimusten materiaaleille asianmukaisen arkistoinnin, joka mahdollistaa tiedon pitkäaikaisen säilytyksen. (Borg ja Kuula 2007).

Digitaalisessa muodossa olevat aineistot saattavat helposti tuhoutua tai muuttua käyttökelvottomiksi. Jotta ne pysyisivät tunnistettavina ja käytettävänä, vaativat digitaaliset aineistot suunnitelmallista aineistohallintaa. Perinteisessä arkistossa paperiaineistojen vääränlainen käsittely tai huonot säilytysolosuhteet saattavat vaurioittaa niitä fyysisesti ja jos paperit erotetaan niiden alkuperäisestä kontekstistaan, niiden alkuperää voi olla hankala jälkikäteen selvittää. Puutteellisesti tehty aineistojen luettelointi ja papereiden laittaminen arkistossa väärin paikkoihin saattaa aiheuttaa sen, että niitä on vaikea löytää. Huolellinen aineistohallinta pienentää riskiä edellä mainittuihin ongelmiin. Kaikki tämä pätee myös digitaalisiin aineistoihin, vaikka niiden muoto ja tavat säilyttää ja hallita niitä ovat erilaisia kuin paperiaineistojen kanssa. (Higgins 2012).

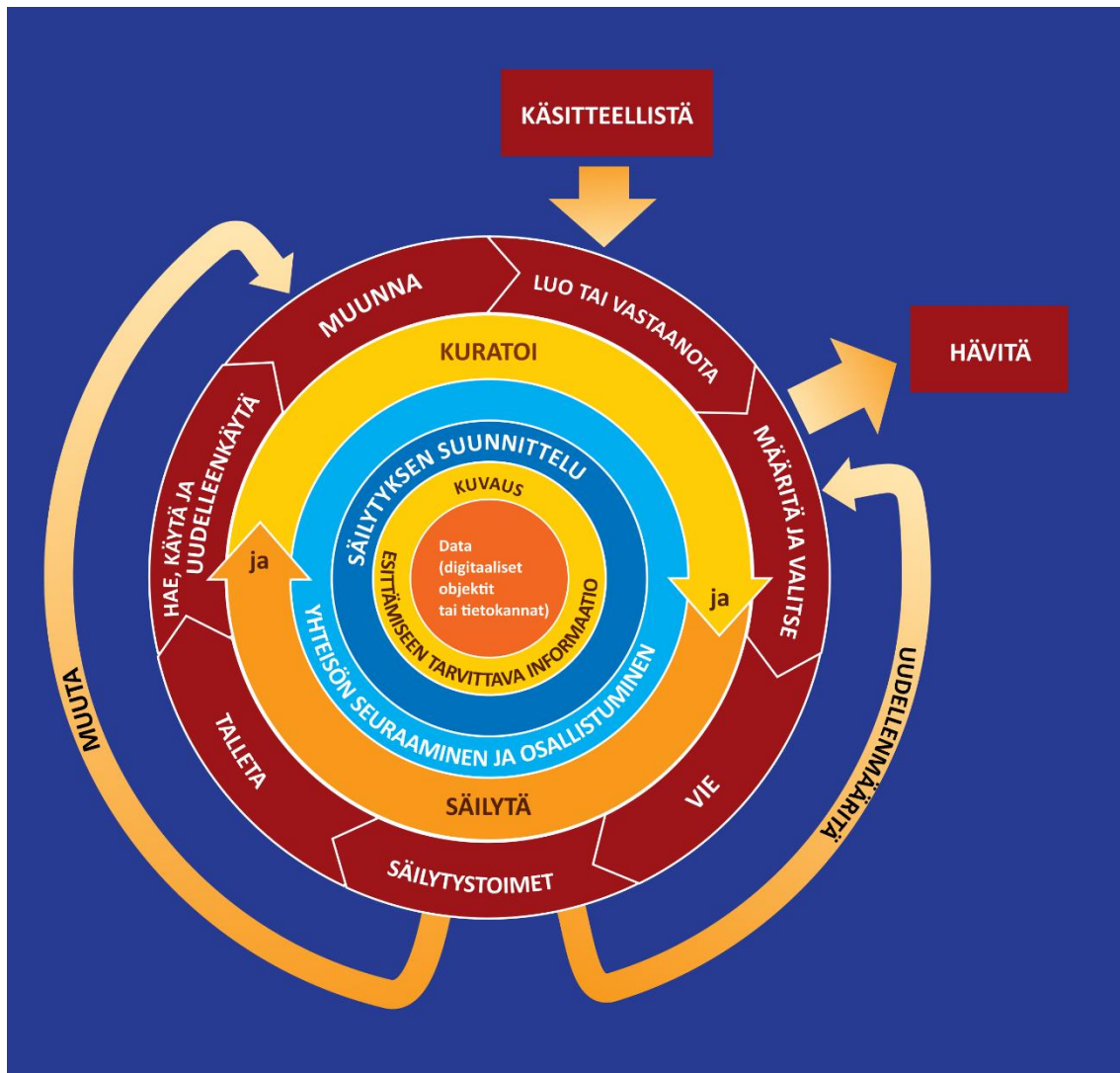
Termiä *digitaalinen kuratointi* (digital curation) on hiljattain alettu käyttää kuvaamaan digitaalisten aineistojen hallintaa tulevaisuuden käyttöä varten. Termiä kuratointi on pitkään käytetty esimerkiksi museoalalla, mutta Yakelin (2007) mukaan itse termi digitaalinen kuratointi on varsin uusi. Digitaalinen kuratointi voidaan määritellä toiminnaksi, jonka tarkoituksena on hallita aineistoa aivan sen syntymästä lähtien ja varmistaa, että aineisto säilyy käytettävänä siihen tarkoitukseen, johon se on luotu, sekä säilyttää se käytettävänä myös myöhempää käyttöä varten. Joidenkin digitaalisten aineistojen suhteen se tarkoittaa jatkuvaa päivittämistä, jotta ne pysyvät käytettävässä muodossa myös tulevaisuudessa. Digitaalisen kuratoinnin tarkoituksena on nimenomaan aineistojen säilyttäminen, jotta niitä voidaan käyttää myös tulevaisuudessa. Digitaalista kuratointia voidaan oikeastaan pitää sateenkaariterminä kuvaamaan digitaalista säilyttämistä ja aineistohallintaa. (Yakel 2007).

Asiakirjahallinnon perinteeseen kuuluvat erilaiset elinkaarimallit. Niiden tarkoituksena on kuvata asiakirjojen elinkaarta sen eri vaiheissa. (Henttonen 2015) Vaikka tutkimusaineistot eroavat esimerkiksi julkishallinnon asiakirjoista ja tutkimusaineistoja arkistivat palvelut eroavat julkishallinnon arkistoista, Pennock (2007) toteaa, että aineistojen elinkaariajattelu olisi hyvä tuoda mukaan myös digitaalisten tutkimusaineistojen pariin. Aineistojen koko elinkaaren hallinnointi olisi ideaalitilanne. Elinkaareen vaiheita tutkimuksessa ovat aineiston luominen, aktiivinen käyttö, arviointi ja valinta, arkistointi ja säilyttäminen, uudelleen käyttö ja saatavuus. Jokaisesta elinkaaren osa-alueesta huolehtiminen vaikuttaa aina seuraavaan osa-alueeseen ja siihen, miten aineistot säilyvät ja säilyttävät autenttisuutensa ja eheydensä. (Pennock 2007).

Perinteinen elinkaarimalli on yksinkertaisuudessaan kuvaus siitä, kuinka asiakirja luodaan, sitä käytetään ja säilytetään ja lopuksi hävitetään tai arkistoidaan. Perinteisessä elinkaarimallissa ajatellaan, että asiakirjat syntyvät organisaatiossa ja niistä huolehtii organisaation asiakirjahallinto ja myöhemmin ne siirtyvät päätearkistoon arkistonhoitajien vastuulle ja nämä kaksi vaihetta ovat eriytetty toisistaan. (Henttonen 2015, 82-83). Sähköisiä aineistoja elinkaarimalli ei välttämättä kuvaa parhaiten, sillä elinkaaren eri vaiheiden erot ovat pienentyneet (Henttonen 2015, 84). Australialainen Frank Upward kehitti 1990-luvun puolivälissä elinkaaren jatkumomallin (records continuum model) kuvaamaan yksityiskohtaisemmin ja kokonaisvaltaisemmin asiakirjojen luonnetta, roolia, käyttöä ja elinkaarta. (Gilliland 2014, 220). Jatkumomallin ideana on, etteivät asiakirjat kulje lineaarisesti eteenpäin kuten perinteisessä elinkaarimallissa. Asiakirjat syntyvät tietyssä tarkoituksessa, mutta niitä voi käyttää myös toisiin tarkoituksiin eri aikoina. Asiakirjan elinkaari ei jakaudu selkeisiin osiin toisin kuin perinteisessä elinkaarimallissa ja asiakirja saattaa olla samaan aikaan kahdessa eri vaiheessa eli sekä aktiivisessa käytössä että osana kulttuurihistoriaa. (Henttonen 2015, 84-85).

Brittiläinen Digital Curation Center on kehittänyt vielä monitahoisemman version elinkaarimallista. *Digitaalisen kuraation malli* (Curation Lifecycle) on suunnattu kaikille aineistojen kanssa tekemisissä oleville, aineistoja luoville, niitä käsitteleville ja käyttäjille, aineistojen tyypistä riippumatta. Sen tarkoituksena on toimia työkaluna, joka auttaa organisoimaan ja suunnittelemaan aineistojen säilyttämistä ja tunnistamaan aineistohallin-

taan liittyviä riskejä. (Gilliland 2014, 223). Malli ei kuvaa prosesseja vaan pyrkii antamaan kuvan siitä, millaiset toiminnot digitaalisten aineistojen kanssa ovat yksittäisiä, mitkä ovat perättäisiä ja mitkä tapahtuvat koko käsiteltävän aineiston elinkaaren ajan. Se myös huomioi sen, että digitaalisia aineistoja voi uudelleen käyttää ja muuntaa edelleen erilaisiin muotoihin. Uudelleen käytön seurauksena puolestaan syntyy uusia aineistoja, joista tulee myös osa elinkaarta. (Gilliland 2014, 224)



Kuva 1

Myös Higgins on kuvannut digitaalisen kuraation mallia. Malli alkaa käsitteellistämisen vaihteesta, jossa työ suunnitellaan siten, että digitaalinen kuraatio otetaan huomioon. Tässä vaiheessa käydään läpi jo aineiston oletettu elinkaari ja määritellään aineiston keräämiseen, käsittelyyn ja säilyttämiseen liittyvät toimenpiteet. Tämän jälkeen aineiston luomisen tai vastaanottamisen vaiheessa keskitytään siihen, että aineisto on käytettyyn

tarkoitukseen tarkoituksen mukaista ja valmis kuratoitavaksi. Seuraavaksi materiaalista valitaan mikä laitetaan pitkäaikaiseen säilytykseen, mitä säilytetään toistaiseksi ja mitä hävitetään. Vaikka digitaalisten aineistojen säilyttäminen on helppoa ja halpaa, aineiston seulonta edistää aineiston pitkäikäisyyttä siten, että vain tärkeä osa aineistosta jää pitkäaikaiseen säilytykseen, pienempää aineistoa on helpompi hallita ja se edistää sitä, että aineisto säilyy käyttökelpoisena pitkään. (Higgins 2012).

Ennen aineiston tallentamista lopulliseen säilytyspaikkaan pitää ottaa huomioon sekä hallinnolliset prosessit että tekniset toimenpiteet, joilla taataan aineiston säilyvyys. Aineiston talletuspaikan ei tarvitse olla virallinen arkisto, vaan se voi olla esimerkiksi instituution oma jaettu palvelintila, kunhan säilytyksen turvallisuus ja aineiston luotettavuus pystytään takaamaan ja aineistoon on pääsy niillä, kenen aineistoa tarvitsee käyttää. Digitaalisen kuratoinnin tavoitteena on, että aineistoja voidaan käyttää uudelleen ja niihin on pääsy myös ensikäytön jälkeen. (Higgins 2012).

2.1.1 Tutkimusaineistojen elinkaari

Seuraavaksi esittelen sähköisen tutkimusaineiston elinkaarta tavanomaisessa tutkimusprosessissa. Aineistohallinnassa tavoitteena on huolehtia tutkimusaineiston hallinnasta jo ennen kuin aineisto on syntynyt aina elinkaaren loppuun asti (Doorn ja Tjalsma 2007). Tutkimusorganisaatiossa aineistohallinnan osa-alueista voidaan vastata kolmella eri tavalla. Yksi tapa on, että tutkijat huolehtivat itse aineistohallinnan eri osa-alueista heti tutkimuksen alusta alkaen ja tutkimuksen päätyttyä tutkijat säilyttävät ja huolehtivat itse aineistoistaan. Toinen tapa on, että tutkijat itse suorittavat aineistohallintaan liittyvät toimenpiteet ja säilyttävät aineistonsa itse, mutta aineistohallinnan ohjauksesta ja suunnittelusta vastaa informaatioalan ammattilainen. Kolmannessa tavassa tutkijat itse huolehtivat aineistohallinnasta tutkimuksen ollessa käynnissä, jonka jälkeen aineisto arkistoidaan ja vastuu aineistosta siirtyy arkistolle. (Higgins 2012).

Suomessa Kansallisarkisto ei ole määrännyt tutkimustoiminnan asiakirjojen säilytysajoista, vaan ne määritellään virasto- ja laitoskohtaisesti (Borg ja Kuula 2007). Esimerkiksi Tampereen yliopiston arkistonmuodostussuunnitelmassa on määritelty, että tutkimusaineistot säilytetään yksiköissä harkinnan mukaan joko pysyvästi tai määräajaksi.

Yliopiston arkiston kanssa voi sopia myös merkittävien aineistojen säilyttämisestä. (Tampereen yliopiston arkistonmuodostussuunnitelma). Tästä poikkeuksena on Suomen Akatemian rahoittamat tutkimushankkeet, joissa Suomen Akatemia suosittelee tutkimusaineistojen arkistointia Yhteiskuntatieteellisen tietoarkistoon (Borg ja Kuula 2007).

Tutkimusaineiston syntymisen ensimmäinen vaihe on tutkimussuunnitelmassa, joka luo pohjan tutkimuksen aloittamiselle ja ohjaa tutkimuksen tekoa. Tutkimussuunnitelmassa esitellään muun muassa tutkimuksen tavoitteet, aineistot ja menetelmät. Tässä vaiheessa on mahdollista myös suunnitella tutkimusaineistojen elinkaarta. Omistus-, hallinta- ja tekijänoikeudet, aineiston tutkimuksen aikainen säilyttäminen ja käyttötavat sekä aineiston arkistointiin ja jatkokäyttöön liittyvät asiat voi suunnitella jo varhaisessa vaiheessa. (Borg ja Kuula 2007). Tämä vaihe on aineistohallinnan kannalta tärkeä, koska suunnitteluvaiheessa tehdyt ratkaisut kulkevat mukana koko tutkimuksen ajan ja vaikuttavat aineiston pitkäaikaiseen säilyvyyteen ja jatkokäyttöön. (Higgins 2012, 23).

Tutkimusta rahoittavat tahot ovat yhä enenevässä määrin alkaneet vaatia tutkijoilta suunnitelmaa aineistohallinnassa. Suomessa esimerkiksi Suomen Akatemia on vuodesta 2008 vaatinut rahoittamiltaan tutkijoilta aineistohallintasuunnitelmaa, joka liitetään osaksi rahoitushakemusta. Aineistohallintasuunnitelmasta tulisi käydä ilmi, miten tutkimusaineistoa kerätään, kuinka kerättyjä aineistoja käytetään, miten niitä säilytetään ja miten mahdollinen jatkokäyttö aiotaan järjestää. (Suomen Akatemia 2015). Suomen Akatemia esittää, että aineistohallintasuunnitelman tekemiselle on useita hyötyjä. Etukäteen suunniteltu aineistojen käyttö edistää tietoaineistojen saatavuutta jatkossa ja tukee tieteen avoimuuden periaatetta. Lisäksi sen tarkoituksena on tehostaa julkisella rahoituksella tehdyn tutkimuksen kustannustehokasta käyttöä ja tutkimuksen tuottaman tiedon hyödyntämistä. Suunnitelman tekeminen lisää myös tietoisuutta tutkimusaineistojen elinkaaresta ja käyttöarvosta. (Laine 2009).

Yhdysvalloissa puolestaan tietoisuus tutkimusaineiston tärkeydestä lisääntyi huomattavasti vuonna 2010, kun suuri rahoittajataho National Science Foundation (NSF) ilmoitti, että rahoitushakemuksiin pitää liittää myös aineistohallintasuunnitelma. Tämän jälkeen myös muut yhdysvaltalaiset rahoittajat alkoivat vaatia aineistohallintasuunnitelmaa ja käytäntö on levinnyt laajasti. (Ray 2014). Tutkimusaineiston elinkaaren toteuttaminen

suunnitelman mukaan ja hyvien tieteellisten käytäntöjen noudattaminen on pääasiassa tutkijan vastuulla (Borg ja Kuula 2007).

Aineistonkeruuvaiheessa tehdyt ratkaisut vaikuttavat pitkälti tutkimusaineiston elinkaareen. Tutkittavia informoidaan aineiston käsittelyyn liittyvistä vaiheista, kuten tutkimuksen tarkoituksesta, luottamuksellisen tiedon suojaamisesta, miten aineistoa käsitellään, kuka aineistoa käyttää ja miten sitä säilytetään ja mahdollisesti hävitetään. Aineiston sisältäessä tunnistetietoja, elinkaareen vaikuttaa se, mitä tutkittavan kanssa on sovittu aineiston käsittelyyn liittyen. (Borg ja Kuula 2007).

Metatiedolla ja aineiston kuvailulla tarkoitetaan tutkimusaineistojen yhteydessä aineistoista kerroittuja kuvailutietoja. Metatiedot voivat kertoa aineiston sisällön sisällön lisäksi myös aineiston kontekstista, eli missä yhteydessä ja mihin tarkoitukseen aineistoi on luotu. Tietoa voi olla myös aineiston rakenteesta, eli aineiston osien rakenteesta ja suhteesta toisiinsa. (Gilliland 2014, 111) (Henttonen 2015, 119). Metatiedoille on olemassa erilaisia malleja, jotka määrittävät missä muodossa ja mitä tietoja aineistosta annetaan ja eri mallien käyttö ja metatietojen määrä riippuu toimintaympäristöstä ja sen asettamista tarpeista sekä lainsäädännöstä. Tutkimusaineistoja vastaanottavat arkistot eivät välttämättä käytä tiettyjä metatietomalleja, mutta metatietoja aineistoilta vaaditaan. Metatiedot ovat edellytys aineistojen käyttökelpoisuudelle, koska ilman tietoja ne ovat vain yksittäisiä tiedostoja, joilla ei ole laajempaa merkitystä. Metatietojen käyttö parantaa myös aineistojen löydettävyyttä. Tutkimusaineistojen kohdalla tärkeitä metatietoja ovat esimerkiksi aineiston alkuperäinen käyttötarkoitus, sen tekijät, tuottajat, aineiston valintaperusteet ja aineistonkeruun tiedot. Lisäksi yksittäiset tiedostot ja muuttujat kuvaillaan ja annetaan aineiston saatavuustiedot, eli missä aineistoa säilytetään ja liittyykö sen käyttöön erityisehtoja. (Yhteiskuntatieteellinen tietoarkisto).

Kerätyn aineiston käyttämistä tutkimussuunnitelmassa esiteltyyn tarkoitukseen sanotaan ensikäytöksi. Tutkimusaineistot kerätään ensisijaisesti ensikäyttöä varten. Aina ei ole täysin selvää, kuka aineiston ensikäytöstä ja mahdollisesta jatkokäyttöön luovuttamisesta vastaa. Aineistoa saattavat kerätä useat eri tutkijat ja tutkimuksen tekoon liittyvät henkilöt. Myös aineiston keräyksen suunnittelussa voi olla osallisena useampi henkilö. (Borg ja Kuula 2007).

Ensikäytön jälkeen on vuorossa aineiston arkistointi. Ensimmäiset aineistoarkistot perustettiin 1960-luvun alussa Amerikassa ja pian tämän jälkeen niitä perustettiin myös Eurooppaan. Aineistoarkistojen tarve lähti siitä, että tutkimuksessa alkoi tulla yhtä enemmän tietokoneelle tallennettua aineistoa. Tiedon säilyttäminen mahdollisti tutkijoille sen, että he pystyivät jälkikäteen tarkastamaan kerätyn aineiston. Alussa aineistoarkistoihin tallennettiin sosiaalitieteiden kyselyaineistoja. Näitä seurasivat 1970-luvun lopulla ja 1980-luvun alussa perustetut digitaalisten tekstien arkistot. 1980- ja 1990-luvun aikana erilaiset aineistoarkistot alkoivat yleistyä. Aineistoarkistoilla ei ollut mitään tekemistä esimerkiksi julkishallinnon arkistojen kanssa, vaan niitä ylläpitivät esimerkiksi yliopistot. (Doorn ja Tjalsma 2007).

Aineistojen säilyvyyden suhteen ei ole välttämätöntä, että aineistoja säilytetään juuri tutkimusaineistoille tarkoitettussa arkistossa, vaan tutkijoilla voi olla myös omia muita säilytysratkaisuja. Pitkäaikaisen säilyvyyden kannalta oleellista on se, että aineistoja aktiivisesti hallitaan, jotta ajan kuluminen ei aiheuta aineiston muuttumista käyttökelvottomaksi. Oleellista on myös, että säilytys on turvattu teknisten ongelmien varalta esimerkiksi varmuuskopioinnin avulla. (Higgins 2012, 35).

Doornin mukaan arkistojen tärkein tehtävä ei ole pelkästään aineiston säilyvyyden turvaaminen vaan myös mahdollistaa aineistojen uudelleenkäyttö (Doorn ja Tjalsma, 2007). Aineiston jatkokäyttö on sitä helpompaa, mitä paremmin aineisto on kuvailtu ja mitä enemmän vaivaa sen arkistointiin on nähty. Arkistoituja aineistoja voi käyttää hyvin monenlaisiin tarkoituksiin, jotka useimmiten liittyvät tutkimukseen tai opetukseen. Aineistoja käytetään esimerkiksi uuden tutkimuksen tausta-aineistona tai aineistosta on voitu asettaa uusia tutkimusasetelmia ja -kysymyksiä, joita ei ole käytetty alkuperäisessä tutkimuksessa. Erityistä arvoa arkistoiduista tutkimusaineistoista saa silloin, kun uuteen tutkimukseen halutaan ajallista tai kansainvälistä vertailua. Arkistoituja tutkimusaineistoja on käytetty myös uusien tutkimusten suunnitteluun sekä tutkimusmenetelmien kehittelyyn. (Borg ja Kuula 2007).

Useat tieteelliset aikakauslehdet ovat alkaneet luoda toimintapoja, jossa artikkeli julkaistaan vasta sen jälkeen, kun tutkimuksessa käytetty aineisto on julkaistu tiedeyhteisön avoimeen käyttöön. Tällä tavoin lukijat pääsevät tarkastelemaan tutkijan esittämiä väitteitä tulosten pohjalta, mikä vähentää virheitä ja jopa vilppiä kuten myös

edistää tutkimustulosten tieteellistä arviointia. (Borg ja Kuula 2007) (Doorn ja Tjalsma 2007).

2.1.2 Aineistohallinnan haasteita

Digitaalisten tutkimusaineistojen on vaikea löytää omaa paikkaansa, koska nykypäivän informaatioteknologia on keskittynyt tiedon nopeaan ja helppoon saatavuuteen, jossa aineiston sijainnilla ja laadulla ei ole suurta merkitystä. Suuri osa uusista Internet-sovelluksista on suunniteltu käsittelemään kaikkein uusinta tietoa eikä sitä, mitä joskus aikaisemmin on tapahtunut. (Doorn ja Tjalsma 2007). Digitaalisia aineistoja on myös helppo siirtää, kopioida, muokata ja poistaa, mikä aiheuttaa haasteita pitää aineistot luotettavina. Lisäksi esimerkiksi metatietojen puuttuminen voi johtaa siihen, että aineistoa ei voida identifioida eikä tämän takia myöskään käyttää. (Higgins 2012).

Tutkimusaineiston pitkäaikaissäilytykseen liittyvät omat tekniset haasteensa, mutta niiden huomioonottaminen ei pelkästään riitä. Digitaalisen tutkimusaineiston säilytyksessä tulee ottaa huomioon myös samat asiat kuin perinteisissä paperiarkistoissa: aineistojen metatiedot, omistajuus, autenttisuus, yksityisyydensuoja ja saavutettavuus. Teemat eivät ole uusia, mutta digitaalinen maailma asettaa sille uudenlaiset haasteet. (Doorn ja Tjalsma 2007).

Lakiin ja tutkimusetiikkaan perustuvat toimenpiteet ovat myös yksi osa aineistohallinnan haasteista. Parry ja Mauthner (2004) käsittelevät kerätyn aineiston omistajasuhteita ja laadullisen aineiston arkistoinnin ja jakamisen haasteita. Kvantitatiivisten aineistojen hallinta ja arkistointi on koettu suhteellisen ongelmattomaksi, mutta kvalitatiivisten aineistojen kanssa tilanne on toisenlainen. Kvantitatiivisia aineistoja onkin arkistoitu ja jatkokäytetty suuremmassa määrin kuin kvalitatiivisia aineistoja. Rahoittajien vaatiessa tutkijoita luovuttamaan myös kvalitatiiviset aineistonsa jatkokäyttöön, käytännön haasteet tulisi tunnistaa ja huomioida. Laadullisessa tutkimuksessa omistajuus, luottamuksellisuus ja anonymiteetti tulee ottaa erityisesti huomioon jatkokäytön osalta. (Parry ja Mauthner 2004).

Tietoarkistot edellyttävät aineistoilta luottamuksellisuutta ja tutkimuksessa mukana olleiden henkilöiden anonymisointia (Yhteiskuntatieteellinen tietoarkisto). Edellytys perustuu

henkilötietolakiin (523/1999). Lain mukaan henkilötiedoilla tarkoitetaan kaikkea luonnollista henkilöä, hänen elinolosuhteitaan tai omaisuuttaan koskevaa tietoa, joka voidaan yhdistää henkilöön tai hänen perheeseensä. Laki ei erikseen määrittele henkilötiedoksi laskettavan tiedon sisältöä tai luonnetta. Henkilötieto voi olla yksityiselämään, ammatilliseen toimintaan, perhe-elämään liittyviä tietoja. Henkilötiedot voivat olla myös mielipiteitä tai väitteitä eikä niiden todennettavuudella tai subjektiivisuudella ole väliä henkilötietolain mukaan. (Henkilötietolaki 523/1999).

Parryn ja Mauthnerin mukaan anonymiteetin varmistaminen saattaa kuitenkin muuttaa aineiston luonnetta tai laatua. Esimerkiksi jos tehdään etnografista tutkimusta jonkin yhteisön toiminnasta, saattaa yksilöiden tunnistaminen olla helppoa, vaikka henkilöllisyydet olisivat salattu. Pienestä yhteisöstä yksilöt voitaisiin tunnistaa pelkkien taustatietojen perusteella. (Parry ja Mauthner 2004).

Toisaalta vaikka tutkimuksessa käytettäisiin suurta joukkoa ihmisiä, tutkimuksen luonne vaikuttaa siihen, kuinka helposti yksilöt voisivat olla tunnistettavissa materiaalista. Tutkittaessa esimerkiksi työn ja perheen yhdistämistä, tutkimusmateriaaliin kysytään paljon henkilökohtaiseen historiaan, perheeseen ja asemaan liittyviä kysymyksiä. Tällöin materiaalista häivytetään yksityiskohtia anonymiteetin turvaamiseksi. Yksityiskohtien hävittäminen puolestaan saattaa vaikuttaa aineiston laatuun ja käytettävyyteen. (Parry ja Mauthner 2004).

2.2 Aiempi tutkimus

Tutkimusaineistojen hallintaan liittyen on tehty tutkimusta sekä Suomessa että kansainvälisesti. Seuraavaksi esittelen muutamia aineistohallinnasta tehtyjä tutkimuksia. Ensimmäiseksi käyn läpi Britanniassa tehtyjä tutkimuksia siitä, voisiko aineistohallinnan käytännöt yhtenäistää maanlaajuisesti. Seuraavaksi kerron tutkimuksista, jotka käsittelevät tutkimusaineiston säilyttämistä ja saatavuutta. Viimeiseksi käyn läpi aineistohallinnan suunnittelusta tehtyjä tutkimuksia.

2.2.1 Aineistohallinnan käytäntöjä Britanniassa

Digitaalisia tutkimusaineistoja syntyy valtavia määriä maailman eri yliopistoissa. Aineistoja on monenlaisessa muodossa ja niitä syntyy hyvin nopeasti. Tämä luo haasteen kehittämään toimintatapoja ja käytäntöjä sekä haasteen ohjata tutkijoita aineistohallintaan liittyvissä asioissa. Haasteista huolimatta tietoisuus tutkimusaineistojen arvosta on kasvussa, koska arkistoidut tutkimusaineistot mahdollistavat tutkimustulosten arvioinnin jälkikäteen, mahdollistavat laajemman aineiston muodostamisen kansainvälisesti eri aineistoista sekä mahdollistavat kerättyjen aineistojen käytön toisiin tutkimuksiin (Pinfield Cox ja Smith 2014).

Sykes (2008) raportoi Britanniassa tehdystä tutkimusprojektista, jonka tarkoituksena oli selvittää, olisiko mahdollista kehittää yhtenäiset, koko Britannian kattavat käytännöt tutkimusaineistojen hallintaan liittyen. Lähtökohtaisesti tutkimusaineistoihin liittyvät haasteet ajateltiin olevan ainakin säilyttämiseen ja saatavuuteen liittyen. Tutkimusaineistot ovat pääosin tutkijoiden tietokoneilla, eivätkä helposti muiden tutkijoiden saatavilla. Puutteelliset metatiedot puolestaan vaikuttavat aineiston uudelleen käyttöön. Lisäksi on huomattu, että tutkijat ovat alkaneet hankkia aineistoilleen säilytyspaikkoja muun muassa erilaisilta palvelimilta, jotka eivät olleet kuitenkaan arkistointiin tarkoitettuja. (Sykes 2008).

Yhtenäisten käytäntöjen tavoitteena on tuoda tutkijoille esiin käytännöllisiä aineistohallinnan käytäntöjä ja ylipäättään keskittyä kehittämään yhtenäisiä toimintatapoja ja standardeja. Yhtenäiset käytännöt mahdollistavat myös aineistojen löytämisen ja tarjoaisivat tutkijoille koulutusta ja välineitä aineistohallintaan. Lisäksi tavoitteena on yhdistää jo olemassa olevat palvelut, tehdä kustannustehokkaita palveluita ja koordinoita aineistopalveluita tarjoavien, aineistoja luovien ja sitä käyttävien yhteisiä käytäntöjä. (Sykes 2008).

Tutkimuksen tulokset osoittivat, että digitaalisia tutkimusaineistoja syntyy yhä kasvavassa määrin hyvin monilta eri tieteenaloilta. Suurin osa aineistoista säilytetään yliopistossa ja 21 % tutkimukseen osallistuneista tutkijoista säilyttää aineistoa kansallisessa tai kansainvälisessä arkistointipalvelussa. Suurin osa tutkijoista jakaa aineistoaan muille tutkijoille, mutta suurin osa jakamisesta on epämuodollista, oman tutkimusryhmän jäsenille

jakamista. Useat tutkijat kuitenkin toivat esiin, että he tarvitsivat enemmän muiden tutkijoiden tuottamia aineistoja käyttöönsä. Ongelmana oli myös yhtenäisten ja johdonmukaisten toimintatapojen puuttuminen. (Sykes 2008).

Tutkimuksessa tuotiin esiin kolme erilaista mahdollista toimintatapaa tutkimuksessa havaitun tilanteen suhteen. Yhtenä vaihtoehtona olisi jättää tilanne entiselleen, mutta se ei ollut toivottu toimintatapa. Toisena vaihtoehtona oli rakentaa täysin keskitetty palvelu, joka koskisi kaikkea kansallista aineistohallintaa, mutta sellaisen rakentaminen olisi monimutkaista ja kallista ja se mitätöisi tähän asti tehdyn työn tutkimusaineistojen säilyttämiseksi. Kolmantena vaihtoehtona, joka suositeltiin toteutettavan, oli luoda koordinoiva palvelu, jonka tarkoituksena oli työskennellä välikätenä tutkimusta rahoittavien tahojen, tutkijoiden ja tutkimusaineistoja arkistovien palveluiden välillä, joka loisi huomattavia etuja aineistohallintaan liittyen. Näitä etuja olisivat esimerkiksi se, että aineiston elinkaaresta huolehditaan hyviä käytäntöjä noudattaen, mikä luo mahdollisuuksia tulevalle tutkimukselle ja vähentää aineistojen tahatonta tuhoutumista. Lisäksi se vähentäisi samasta aiheesta kerättyjen tutkimusaineistojen määrää. (Sykes 2008).

Cox ja Pinfield (2014) ovat tutkineet tieteellisten kirjastojen roolia tutkimusaineistojen hallinnassa Britanniassa. Jo aiemmin aiheesta on tehty tutkimusta ja niissä todettu, että kirjastojen toimenkuvaan sopisi hyvin myös toimia tutkijoita tukevana tahona, josta saisi neuvoa muun muassa aineistohallintaan, vaikkei kirjaston rooli tutkimusaineistojen hallintaa tukevana tahona aivan ongelmaton ole. (Cox ja Pinfield 2014).

Tutkimuksesta kävi ilmi, että nykytilanteessa Britanniassa tieteelliset kirjastot tarjoavat jonkin verran tutkimusaineistojen hallintaan liittyviä palveluita. Kirjastot itse näkivät tutkimusaineistojen hallinnan tärkeänä toimintana, jossa kirjastot voisivat ottaa suuremman roolin. Kirjastojen asenteet aineistohallintatehtäviä kohtaan olivat muuttuneet lyhyessä ajassa, sillä aiemmissa tutkimuksissa kirjastot kokivat aineistohallintapalvelut enemmän epärelevantteina omalle toiminnalleen kuin nykytilanteessa. Tutkimuksen mukaan kirjastot Britanniassa vaikuttavat olevan myös edelläkävijöitä, mitä tulee palveluiden tarjoamiseen ja uusien toimintatapojen kehittämiseen. Kirjastot ovat antaneet yhä suuremmassa määrin apua esimerkiksi tutkijoiden aineistohallintasuunnitelmien tekoon. Tämän kaltaisen aineistohallintaan liittyvä neuvontatyö koettiin sopivan hyvin kirjastojen toimenkuvaan. (Cox ja Pinfield 2014).

2.2.2 Tilanne Suomessa

Borgin ja Kuulan tutkimusraportissa selvitettiin OECD-aineistosuosittelusta ja pohjustettiin sen kansallista toimeenpanoa. Suosituksen mukaan tutkimusaineistojen avoin saatavuus on tärkeää, jotta julkisrahoitteinen tutkimus voi olla tehokasta, taloudellista sekä olla tuloksellista myös kansainvälisestä näkökulmasta. Suosituksen tarkoituksena on sähköisten tutkimusaineistojen jatkokäytön lisääminen sen jälkeen, kun aineiston alun perin koonneet eivät enää hyödynnä sitä omassa tutkimuksessaan. Avoimuuden hyödyt liittyvät myös tutkimuksen laadunvarmistamiseen. Tutkimusaineistojen avoimuus tukee tieteen avoimuutta ja tutkimustulosten toistettavuutta. (Borg ja Kuula 2007)

He teettivät verkkokyselyn suomalaisten yliopistojen professoreille ja tutkijainstituuttien henkilökunnalle. Vastaajat olivat humanististen-, yhteiskunta- ja kasvatustieteiden aloilta. Kyselyssä selvitettiin muun muassa tutkimusaineistojen säilyttämistä, avointa saatavuutta ja jatkokäyttöä.

Kyselyn perusteella tilanne vuonna 2008 oli se, että suurin osa vastaajista säilytti keräämänsä tutkimusaineistonsa itsellään ja silloin pääosin vain tutkijalla itsellään on mahdollisuus käyttää aineistoja. Vastaajista 29 % kertoi aineistoa säilytettävän omassa yksikössä tai laitoksessa, mutta aineistoa ei ole luetteloitu tai kuvailtu. Vastaajista 12 % kertoi, että aineisto on arkistoitu yksikköön ja asianmukaisesti kuvailtu luetteloihin tai tietokantoihin. Vain 10 % vastaajista kertoi arkistoivansa aineistonsa yliopistosta erilliseen, tutkimusaineistoja vastaanottavaan arkistoon.

Suomessa kerätään valtavia määriä tutkimusaineistoja yliopistoissa ja valtioin tutkimuskeskuksissa, mutta vain pieni osa siitä on arkistoitu ja uudelleen käytetty. Jos tutkimusaineistoja säilytetään tutkimusyksikössä, ei niiden tallennusmuotoa välttämättä pidetä ajan tasalla ja ne voivat muuttuvat käyttökelvottomiksi. Yliopistojen kohdalla kyselyn perusteella tilanne on se, että yliopistolla itsellään ei ole kattavia ohjeistuksia tutkimusaineistojen arkistointiin liittyen eikä niin tueta avoimuutta. Yliopistojen arkistonmuodostussuunnitelmat määrittävät asiakirjojen käsittelyä ja säilyttämistä, mutta ne kattavat lähinnä hallinnon asiakirjat. Yksiköillä saattaa olla ohjeistuksia tutkimusaineistojen hallintaan liittyen ja esimerkiksi Yhteiskuntatieteellisen tietoarkiston vaikutus näkyy varsinkin yhteiskuntatieteellisen tutkimuksen arkistoinnissa. Yliopiston puutteellinen ohjeistus on

kuitenkin ongelmallisista, sillä yliopistojen tulisi säilyttää julkisin varoin toteutetuissa tutkimuksissa syntynyt aineisto. Aineistojen kokonaishallinnan sijaan vastuu aineistosta on jäänyt laitostasolle ja yksittäisten tutkimusryhmien tai jopa tutkijoiden tehtäväksi. (Borg ja Kuula 2007).

Tutkijoilta itseltään kysyttäessä havaittiin, että jatkokäytölle oli useita esteitä. Tutkijoista lähes joka toinen piti ongelmana sitä, etteivät muut osaisi käyttää aineistoa oikealla tavalla tai etteivät muut voi ymmärtää aineistoa. Lisäksi kaksi kolmasosaa piti toisten ko-koamia aineistoja soveltumattomina omiin tutkimuksiinsa. (Borg ja Kuula 2007).

Epäselvyyttä tutkijoiden keskuudessa aiheutti myös tekijänoikeuslainsäädäntö ja se, miten se tutkimusaineistoihin vaikuttaa. Yleensä tutkimusaineiston kohdalla teoskynnys ei ylity eli niitä ei ole suojattu tekijänoikeudella. Asiaa monimutkaistaa se, että tekijänoikeuslainsäädännössä säädetään eräistä tekijänoikeutta lähellä olevista oikeuksista, joiden mukaan esimerkiksi tietokanta tai luettelo voi saada 15 vuoden suoja-ajan. Myös henkilötietolaki luo esteitä. Henkilötietoja saa käsitellä, jos siihen on saatu tutkittavan vapaaehtoinen, yksilöity ja tietoinen tahdonilmaisu. Lainsäädännön epäselvyyden lisäksi jatkokäyttöä vaikeuttivat epäviralliset akateemiset käytännöt ja ajattelutavat. Koska aineiston käsittely vaatii tutkijoilta paljon työtä ja resursseja, voivat tutkijat ajatella, etteivät halua jakaa aineistoaan muiden käytettäväksi. Tutkijoita vaivaa siis mustasukkainen suhtautuminen aineistoonsa. (Borg ja Kuula 2007).

Jatkokäytön esteenä nähtiin myös se, ettei aineiston jatkokäyttöä tai arkistointia oteta huomioon aineistoa kerätessä. Aineistojen sisällöt ovat puutteellisesti dokumentoituja, ja ainoastaan aineiston koonnetut tutkijat osaavat tulkita aineistoa. Tilannetta voitaisiin parantaa esimerkiksi tutkimuksen ja koulutuksen kautta. (Borg ja Kuula 2007).

2.2.3 Aineistohallinnan suunnitleminen

Dietrich, Adamus, Miner ja Steinhart (2012) tutkivat amerikkalaisten tutkimuksia rahoittavien tahojen käytäntöjä aineistohallinnan ohjeistuksen suhteen. Tieteelliset kirjastot ovat alkaneet tarjota apua tutkijoille aineistohallintasuunnitelmien tekoon ja on tärkeää, että koulutusta tarjoavalla taholla on parhaat valmiudet tukea tutkijoita tekemään tar-

peeksi kattavia suunnitelmia. Tutkimusta rahoittavat tahot haluavat rahoittamistaan tutkimuksista aineistohallintasuunnitelman ja tutkimuksen tavoitteena oli selvittää, löytyykö suunnitelmaan teosta ja aineistohallinnan käytännöistä sellaisia ohjeistuksia, joiden avulla tutkijat voivat tehdä vaatimukset täyttäviä suunnitelmia.

Tutkijat valitsivat kymmenen rahoittajaa ja tutkivat rahoittajien verkkosivuilta löytyviä ohjeistuksia aineistohallinnan käytännöistä. Löydettyjä ohjeistuksia analysoitiin 17 kriteerin avulla, joiden pääteemat olivat yleiset käytännöt, standardit, aineiston saatavuus ja säilyttäminen sekä tutkimusjulkaisut.

Tuloksena saatiin, että tutkimuksessa asetetut analysointikriteerit täyttyivät melko puutteellisesti. Yhdelläkään rahoittajista ei löytynyt täysin kattavia ohjeistuksia aineistohallinnan käytännöistä ja eikä neljältä rahoittajalta löytynyt minkäänlaisia ohjeita aineistohallinnan käytäntöihin.

Rahoittavat tahot vaativat yhä enenevässä määrin suunnitelmaa tutkimusaineistojen arkistointiin ja jatkokäyttöön liittyen. Tutkimuksessa todettiin, että eri tahojen periaatteet tutkimusaineistojen suhteen vaihtelivat ja ohjeistuksissa oli eroja. Ongelmana oli, että aineistohallinnan tavoitteet olivat kyllä selvillä, mutta sen toteutusta ei ollut tarpeeksi yksityiskohtaisesti selvitetty. Tutkijoille hankaluuksia aiheuttaa se, jos ei tiedä keinoja oman aineistonsa jakamiseen ja säilyttämiseen ja toisaalta myös se, jos ei saa rahoitusta aineistohallinnan mahdollisiin kuluihin. (Dietrich ym. 2012)

Steinhart, Chen, Arguillas, Dietrich, ja Kramer (2012) tutkivat tutkijoiden valmiuksia tehdä rahoittajien vaatimusten mukaisia aineistohallintasuunnitelmia. Tutkimuksen tarkoituksena oli parantaa yliopiston tarjoamia palveluita tutkijoille. Tutkimus toteutettiin verkkokyselyllä, jossa oli 43 kysymystä. Kyselyssä selvitettiin, minkä tyyppistä materiaalia tutkimuksessa syntyy ja mitkä ovat tutkijoiden tavat käsitellä aineistoja ja metatietoja. Tutkijoille esitettiin myös kysymyksiä tutkijoiden käytännöistä ja ehdoista tutkimusaineiston saatavuuden, jakamisen, luotettavuuden, tietoturvan ja omistajuuden suhteen. Lisäksi kysymyksiä oli myös jatkokäytöstä ja arkistoinnista. Kysymykset perustuivat yhdysvaltalaisen National Science Foundation – rahoittajan vaatimukseen aineistohallinta-

suunnitelman sisällöstä. Kyselyyn saatiin 86 vastausta ja vastausprosentti oli 5,2. Kaikkiin kysymyksiin ei ollut pakko vastata, mutta vain kaksi vastaajaa vastasi alle 20 kysymykseen 43:sta kysymyksestä. (Steinhart ym. 2012)

Tutkimuksen tuloksena oli, että tutkijat olivat vielä epävarmoja siitä, miten rahoittajien vaatimuksiin pitäisi vastata ja millaisia suunnitelmia tulisi tehdä. Suurin osa vastaajista toivoi saavansa apua aineistohallintasuunnitelman tekemiseen. Epävarmuutta aiheuttivat esimerkiksi kokemuksen puute aineistohallinnan eri vaiheisiin liittyvistä toimenpiteistä. (Steinhart ym. 2012)

Kyselyssä selvisi, että tutkijat tuottavat hyvin eri tyyppistä materiaalia ja monissa eri tiedostomuodoissa. Ongelmia aiheutti muun muassa se, että joillain tutkijoilla oli vaikeuksia tunnistaa, mikä syntyvästä materiaalista on itseasiassa tutkimusaineistoa ja mikä siitä on tarpeen säilyttää. Lisäksi Yhdysvalloissa on monia aineistopalveluita tarjoavia yrityksiä ja he tarjoavat säilytystä hyvin laajalle valikoimalle erilaisia digitaalisen materiaalin muotoja. Tämä puolestaan lisää haastetta niille tahoille, joiden tarkoituksena on saada tutkimusaineistot säilymään pitkään. (Steinhart ym. 2012)

Kyselyn perusteella suurin osa tutkijoista oli valmiita jossain määrin antamaan aineistonsa jatkokäyttöön. Yleisimmät syyt olla antamatta aineistoa olivat luottamuksellisuuden ja yksityisyyteen liittyvät syyt sekä se, että tutkijat kokivat, etteivät heidän aineistonsa toisi hyötyä muille tutkijoille. Kyselyssä selvisi, että tutkijat tarvitsisivat ohjausta siihen, minkälaisen aineiston jakamisesta olisi hyötyä muille tutkijoille ja mitkä ovat ne luottamukseen ja yksityisyyteen liittyvät tekijät, jotka saattavat estää aineiston jakamisen.

Yhä useammat tieteelliset lehdet ovat muodostaneet uusia käytäntöjä aineistojen saatavuuden suhteen. Tarkoituksena on, että lukijoilla olisi mahdollisuus tarkastaa tutkijan tekemät väitteet tutkimuksessa käytetyn aineiston perusteella. (Doorn ja Tjalsma 2007) Vines (2013) tutki yhdessä muiden tutkijoiden kanssa arkistointivaatimusten vaikutusta arkistoitavien tutkimusaineistojen määrään. Kyseessä oli tieteellisten lehtien erilaiset vaatimukset julkaistavien artikkeleiden tutkimusaineistoihin liittyen. Artikkelin mukaan tutkimusjulkaisujen taustalla olevien tutkimusaineistojen säilyttäminen on tärkeää osaa tutkimusprosessia ja oleellista tutkimuksen toistettavuuden kannalta. Kaikissa tapauksissa tutkimusaineistot eivät ole kuitenkaan saatavissa. (Vines ym. 2013)

Tarkastelun kohteeksi otettiin neljä erilaista tieteellisissä lehdissä esiintyvää arkistoinnin vaatimustasoa ja katsottiin, miten nämä erilaiset vaatimukset vaikuttivat arkistoitavien tutkimusaineistojen määrään. Vaatimustasot olivat: ei vaatimuksia arkistointiin, suositellaan arkistointia ja kaksi kahdessa tapauksessa vaadittiin arkistointia hieman eri perustein. Toisessa tapauksessa vaadittiin arkistointia johonkin julkiseen arkistoon, kun taas toisessa tapauksessa tutkijan piti liittää artikkelin yhteyteen lausunto siitä, että tutkimusaineisto on saatavilla ja missä (tietokannassa) se sijaitsee. Tutkimusaineisto valittiin vuosina 2011 ja 2012 julkaistuista artikkeleista. Tutkimukseen valittiin tieteellisiä lehtiä jokaisesta neljästä vaatimuskategoriasta ja tutkimuksen ulkopuolelle jätettiin ne lehdet, joissa oli alle viisi tutkimukseen soveliaista artikkelia.

Tutkimuksesta selviää, että lehtien arkistointivaatimuksilla oli suuri merkitys tutkimusaineistojen arkistointiin. Tutkimusaineistossa oli 88 artikkelia, jotka oli julkaistu lehdissä, joilla ei ollut arkistointivaatimuksia. Näistä ainoastaan kolmessa tapauksessa tutkimusaineisto oli saatavilla. Suositeltu tutkimusaineistojen arkistointi paransi tilannetta hieman. Kun artikkeleja oli 79, oli kymmenen artikkelin tutkimusaineistot arkistoitu ja saatavilla. Vaatimus tutkimusaineistojen arkistoinnista lisäsi arkistoitujen tutkimusaineistojen määrää. Niissä tapauksissa, joissa vaadittiin arkistointia, mutta ei lausuntoa, kuudessa tapauksessa kuudestatoista aineistot arkistoitiin. Lausunto arkistoinnista paransi tilannetta entisestään ja tutkimusaineistot arkistoitiin 34 tapauksessa 35:stä.

3 TUTKIMUSMENETELMÄT

Tässä luvussa käyn läpi tutkimuskysymykset, tutkimukseen kerättävän aineiston ja tutkimusmenetelmät. Luvussa 3.1 esittelen tutkimuskysymykset, joiden tarkoituksena on ottaa selvää Tampereen yliopiston tutkijoiden aineistohallintatavoista, suhtautumisesta aineistohallintaan sekä mahdollisista aineistohallinnan haasteista. Luvussa 3.2 kerron tutkimuksessa käyttämästäni tutkimusmenetelmästä, teemahaastattelusta. Viimeiseksi luvussa 3.3 kerron aineiston keräämisestä ja siihen vaikuttaneista tekijöistä.

3.1 Tutkimuskysymykset

Olen asettanut tutkimukselleni kaksi keskeistä tutkimuskysymystä. Tavoitteenani on selvittää Tampereen yliopiston tutkijoiden näkemyksiä ja kokemuksia tutkimusaineistojen hallintaan liittyvissä tavoissa ja haasteissa.

Tutkimuskysymykseni ovat:

- Minkälaisia osa-alueita tutkijat tunnistavat aineistohallinnassa ja minkälaisia tapoja tutkijoilla on tehdä aineistohallintaa?
- Mitä ongelmia tutkijat kohtaavat aineistohallinnassa?

Ensimmäisen tutkimuskysymyksen tarkoituksena on selvittää tutkijoiden omaa tapaa tehdä aineistohallintaa. Haluan tietää, minkälaisia aineistohallinnan osa-alueita tutkijat tunnistavat ja kuinka järjestelmällisesti he itse oman tutkimusaineistonsa kanssa tekevät aineistohallintaa.

Toisen tutkimuskysymyksen tarkoituksena on selvittää, ovatko tutkijat kohdanneet aineistohallintaan liittyviä ongelmia ja jos ovat, niin minkälaisia ongelmat ovat.

3.2 Tutkimusaineisto ja –menetelmät

Tutkielmani tarkoituksena on selvittää aineistohallintaa tutkijoiden näkökulmasta. Ihmisten toimintaa ja näkemyksiä kuvattaessa laadullinen tutkimusote sopii tiedon keruun

välineeksi. Laadullisia tutkimusmetodeja suositaan usein silloin, kun halutaan saada tutkittavien ääni esille. (Hirsjärvi, Remes, ja Sajavaara 2009, 164) Tutkielmani on laadullinen tutkimus, aineistonkeruumenetelmänä on käytetty teemahaastattelua ja analysointimenetelmänä on sisällönanalyysi.

3.2.1 Laadullinen tutkimus ja teemahaastattelu

Päädyin laadulliseen tutkimusmenetelmään koska tavoitteenani on saada yleiskuvaa tutkijoiden aineistohallintaan liittyvistä tavoista, näkemyksistä ja haasteista. Laadullisen tutkimuksen tarkoituksen on kuvata todellisia ja luonnollisia tilanteita. Todelliset ja luonnolliset tilanteet ovat monitahoisia ja monet eri tekijät vaikuttavat toisiinsa ja laadullisen tutkimuksen tavoitteena on kuvata tilanteita mahdollisimman kokonaisvaltaisesti. (Hirsjärvi ym. 2009, 162)

Tutkimus toteutetaan haastattelumenetelmällä. Tutkimuksen olisi voinut toteuttaa myös esimerkiksi kyselytutkimuksena, mutta ajattelin haastattelumenetelmän antavan yksityiskohtaisempaa tietoa tutkimastani aiheesta sekä pelkäsin myös kyselytutkimuksen vastausmäärän jäävän liian pieneksi. Tutustuessani aiempaan tutkimukseen aineistohallinnasta, tutustuin myös tutkimukseen, jossa oli käytetty tutkimusmenetelmänä verkkokyselyä. Kyselyn vastausprosentti oli melko pieni ja osalla vastaajista oli ongelmana se, että kysymyksiin vastaaminen oli hankalaa, koska he eivät tienneet aiheesta tarpeeksi tai eivät välttämättä ymmärtäneet kysymystä. Tästä syystä valitsin laadullisen tutkimusotteen ja haastattelun metodikseni, jonka etuina ovat joustavuus sekä mahdollisuus toistaa kysymyksiä, oikaista väärinkäsityksiä sekä käydä keskustela haastateltavan kanssa (Tuomi ja Sarajärvi 2009, 73).

Tutkimusta tehdessä haastattelu on joustava menetelmä ja se sopii monenlaisiin tutkimustarkoituksiin. Haastattelutilanteessa tutkija on suorassa vuorovaikutuksessa tutkittavan kanssa, ja tämä antaa mahdollisuuden viedä haastattelua haluttuun suuntaan ja selvittää motiiveja vastausten taustalla. (Hirsjärvi ja Hurme 2008, 34) Laadulliselle tutkimukselle tyypillistä on myös se, että tutkimuksen kohdejoukko valitaan tarkoituksen mukaisesti (Hirsjärvi ym. 2009, 164) ja tutkittavien tapausten määrä pidetään varsin pienenä (Eskola ja Suoranta 1998, 18).

Tutkimushaastattelun lajeja on useampia. Ne eroavat toisistaan lähinnä sen perusteella, kuinka strukturoituja ne ovat eli kuinka kiinteitä haastattelukysymyksiä tutkija on tehnyt ja miten haastattelutilanne on jäsennetty. (Hirsjärvi ja Hurme 2008, 43) Omassa tutkielmassani käytän puolistrukturoitua haastattelumenetelmää eli teemahaastattelua. Teemahaastattelu on puolistrukturoitu siitä syystä, että haastattelun teemat ovat kaikille tutkitaville samat, vaikka kysymysten asettelu ja järjestys vaihtelisivat. (Hirsjärvi ja Hurme 2008, 48)

Päädyin valitsemaan teemahaastattelun tutkimusmetodikseni sen joustavuuden vuoksi. Teemahaastattelun haasteena on se, että haastattelussa saattaa saada tutkimuksen kannalta paljon epärelevanttia materiaalia (Hirsjärvi ja Hurme 2008, 36). Menetelmässä on kuitenkin tutkimusaiheeni kannalta erityisiä etuja. Haastattelemillani tutkijoilla saattaa olla eritasoista tuntemusta ja kokemusta aineistohallinnasta, jolloin kiinteisiin haastattelukysymyksiin vastaaminen ei välttämättä olisi mielekäästä. Haastattelun kuluessa saatujen vastausten perusteella voi tarkentaa tiettyihin teemoihin tai kysyä vähemmän jostakin teemasta. Mahdollisuus selventää ja syventää haastattelussa vastauksia pyytämällä esimerkiksi mielipiteiden perustelua tai esittämällä lisäkysymyksiä on etu aineiston keräämisen kannalta tutkimassani aiheessa. (Hirsjärvi 2008).

Teemahaastattelussani on kaksi teemaa: toimintatavat sekä ongelmat. Molemmat teemat vastaavat yhtä tutkimuskysymyksistäni. Kysymykset olivat:

- Minkälaista aineistoa tutkimuksessasi syntyy?
- Onko ennen aineistonkeruun aloittamista tehty aineistohallintasuunnitelmaa?
 - o Jos kyllä, millaisia asioita siinä on tuotu esiin?
 - o Onko etukäteen tehty suunnitelma auttanut aineistohallintaa tutkimuksen aikana?
 - o Onko aineiston koko elinkaari pystytty ottamaan huomioon suunnitelmaa tehdessä?
- Voitko kuvailla aineistonkeruuprosessia?
- Miten aineistoa säilytetään ja käsitellään tutkimuksen aikana?
- Miten kerätty aineisto on turvattu katoamiselta tai sekaantumiselta?
- Mitä aineistolle tapahtuu tutkimuksen loputtua? Arkistoidaanko se?

- Mitkä asiat vaikuttavat aineiston arkistointiin tai arkistoimatta jättämiseen?
- Voiko sitä käyttää uudelleen muissa tutkimuksissa? Onko siihen liittyen rajoituksia?
- Kuinka järjestelmällistä aineistohallinta on?
- Millä tavoin aineistohallinnan kysymykset nousevat esiin työssä tai tutkimuksessa?
- Oletko kohdannut ongelmia aineistohallintaan liittyen? Minkälaisia?
- Mistä saa apua aineistohallinnan eri ongelmiin?

3.2.2 Aineiston kerääminen

Haastattelin seitsemää Tampereen yliopiston tutkijaa. Tarkoitukseni oli saada haastateltavaksi tutkijoita eri yksiköistä sillä perusteella, että haastateltavien tutkimusaineistot ovat erilaisia ja tutkimukseen saataisiin sekä kvalitatiivisen että kvantitatiivisen tutkimuksen edustajia. Lähetin haastattelupyynnöjä eri yksiköiden tutkijoille. Valitsin sattumanvaraisesti usean eri yksikön yhteystiedoista tutkijoita, joille lähetin haastattelupyynnön.

Haastattelut suoritin toukokuussa ja kesäkuussa 2016. Haastatteluiden kesto vaihteli 20 minuutin ja 50 minuutin välillä. En tehnyt pilottihaastattelua ennen varsinaisia tutkimushaastatteluita. Jälkikäteen ajatellen pilottihaastattelun tekeminen olisi voinut olla järkevää varsinkin ensimmäisten haastatteluiden sujuvuutta ajatellen.

Haastateltavaksi sain seitsemän tutkijaa neljästä eri yksiköstä: yhteiskuntatieteiden yksiköstä, terveystieteiden yksiköstä, johtamiskorkeakoulusta ja BioMediTech:stä. Yhteiskuntatieteiden yksiköstä haastattelin kolmea eri alan kvalitatiivista tutkimusta tekevää tutkijaa. Johtamiskorkeakoulusta haastattelemani tutkija teki myös kvalitatiivista tutkimusta. Terveystieteiden yksiköstä haastattelin kahta samantyyppisissä tutkimuksissa mukana olevaa tutkijaa. He eivät ole varsinaisia tutkimuksen toteuttajia, vaan tutkimusryhmässä mukana olevia, jotka huolehtivat aineistohallinnan eri osa-alueista ja ohjaavat tut-

kijoita toimimaan oikein. BioMediTech:stä haastattelin yhtä tutkijaa, jonka tutkimusaineistona olivat sekä solunäytteet että sähköiset ja paperiset tutkimusaineistot. Tutkielmasani viittaa tämän tutkijan kohdalla näihin sähköisiin ja paperisiin tutkimusaineistoihin.

Aineiston keräämisen jälkeen litteroin kirjakielelle muuttamatta haastattelujen sanomaa. Tarkoitukseni ei ollut selvittää, miten asiat kerrottiin vaan mitä kerrottiin. Lainaukset eivät ole sananmukaisia. Litteroinnin jälkeen aloitin aineiston analyysin, josta kerron seuraavassa luvussa.

3.2.3 Aineiston analyysi

Aineiston analyysin menetelmänä käytin sisällönanalyysiä. Sisällönanalyysi on perinteinen menetelmä laadullisen tutkimuksen analysoinnissa ja sitä voidaan käyttää monenlaisessa tutkimuksessa (Tuomi ja Sarajärvi 2009, 91). Sisällönanalyysillä pyritään samaan tutkittavasta ilmiöstä tiivis kuvaus ja se on väline, jolla kerättyyn aineistoon saadaan järjestys johtopäätösten tekoa varten. Sisällönanalyysi on tekstianalyysiä, jonka avulla tekstistä etsitään merkityksiä.

Kävin läpi aineistoni siten, että litteroiduista aineistoista poimin ne asiat, jotka ovat oleellisia tutkimukseni kannalta. Tämän jälkeen luokittelin aineistoni tutkimuskysymysteni teemojen mukaan, jonka jälkeen vielä jaoin aineistoin pienempien aihekokonaisuuksien perusteella. Tämän avulla sain jäsenneltyä haastatteluista tutkimuksen kannalta oleelliset asiat ja kirjoitettua ne auki seuraavaan lukuun 4, jossa esittelen tutkimustulokset.

4 TUTKIMUSTULOKSET

Tässä luvussa käyn läpi haastatteluista saamiani tuloksia. Ensimmäisessä alaluvussa käyn läpi tutkijoiden kokemuksia ja käytäntöjä aineistohallintaan liittyen. Toisessa alaluvussa tuodaan esiin minkälaisiin ongelmiin tutkijat ovat törmänneet aineistohallinnassa. Laadullista tutkimusta tekeviin tutkijoihin viitataan koodeilla L1, L2, L3, L4 ja määrällistä tutkimusta tekeviin koodeilla M1, M2 ja M3.

4.1 Aineistohallinnan käytännöistä

Tässä alaluvussa käyn läpi tutkijoiden kokemuksia aineistohallinnasta. Käyn läpi, miten he hallitsivat aineistoa ennen tutkimuksen alkua, tutkimuksen aikana ja tutkimuksen päätyttyä. Ensin kerron tulokset aineistohallinnan suunnittelusta, tämän jälkeen aineistohallinnasta tutkimuksen ollessa käynnissä ja viimeiseksi aineiston mahdollisesta arkistoinnista ja jatkokäytöstä.

4.1.1 Suunnittelusta

Tutkijoiden kanssa keskusteltiin siitä, ovatko he suunnitelleet aineiston hallintaa jo ennen tutkimuksen alkamista tai aivan tutkimuksen alkumetreillä. Keskusteluissa käytiin myös läpi etukäteissuunnittelun tarpeellisuutta ja onko mahdollisesta suunnittelusta ollut hyötyä tutkimuksen edetessä. Esiin nousi pohdintaa myös siitä, onko aineistohallinnan suunnittelu etukäteen edes mahdollista ja kuinka helposti suunnitelmat muuttuvat käytännöiksi.

Laadullisen tutkimuksen suunnittelussa pääpaino oli pitkälti aineistojen toimittamisessa arkistoon. Kvantitatiivisen tutkimuksen suunnittelussa keskityttiin paljon aineistohallintaan tutkimuksen aikana, kuten lakeihin ja yksityisyyden suojaan.

Laadullisen tutkimuksen edustajista suurin osa on suunnitellut aineistohallintaa jo ennen tutkimuksen alkua (L1, L2, L3). Suunnitelmaan on kuulunut esimerkiksi tieto siitä, miten aineistoa kerätään, mihin sitä käytetään, miten sitä käsitellään tutkimuksen aikana ja mihin aineistot päätyvät tutkimuksen loputtua. Yksi vaikuttaja laadullisen aineiston hallinnan suunnittelulle on ollut se, että rahoitushakemukseen tarvitaan tutkimussuunnitelman

liitteeksi myös suunnitelma aineistohallinnasta. Haastatteluissa ilmeni, että oleellinen ja merkittävä osa suunnitelmaa on se, tallennetaanko aineistot jatkokäyttöä varten tutkimuksen loputtua.

Aineistohallinnan suunnittelusta ennen tutkimuksen alkua nousi esiin erilaisia toimintatapoja sekä mielipiteitä. Keskustelin tutkijoiden kanssa siitä, kokevatko he suunnittelun helpoksi vai vaikeaksi ja onko suunnittelusta jotain hyötyä tutkimuksen aikana ja sen päätyttyä.

Yksi haastattemistani tutkijoista kertoi, että aineistohallinnan suunnittelu ei tuota hänelle vaikeuksia, vaan se sujuu pikemminkin luonnostaan. Hän on yrittänyt ajatella elinkaarimallin mukaisesti ja pitänyt aineistoja jo pidemmän aikaa tärkeinä ja huolehtimisen arvoisina.

Silloin, kun minä tein väitöskirjan, oli se tällainen haastatteluihin perustuva hanke, ja siinä yhteydessä jo pyysin haastatelluilta luvan, että nämä saa arkistoida. Olen pitkään ollut sitä mieltä, että ne ovat arvokkaita ja niitä ei saa heittää roskeen sen jälkeen, kun tutkimus on tehty, vaan niitä täytyy huoltaa ja niistä täytyy pitää huolta, että niitä voi käyttää toisenlaisen tutkimuksen tekemiseen. Koska tämä ajatus on minulla aina mielessäni, niin sen auki kirjoittaminen suunnitelmaan ei ole minulle henkilökohtaisesti kauhean vaikeaa. (L3)

Toisenlaisiakin suhtautumisia suunnitelman tekemiseen nousi esiin. Sitä pidettiin ehkä enemmänkin yhtenä pakollisena työvaiheena, kuin hyödyllisenä työkaluna.

Se on vähän niin, että jos jotain pitää sanoa, niin olen yleensä sitten sanonut, että nämä arkistoidaan. Meillä on sellainen aika aktiivinen tutkimusryhmä, niin olen sanonut, että nämä arkistoidaan tutkimusryhmän arkistoon, mikä käytännössä tarkoittaa tuota kovalevyä tuossa minun takanani. Ja sitten, että mahdollisuuksien mukaan tietoarkistoon. (L1)

Vaikka itse suunnitelman tekeminen olisi ollut yksinkertaista, saattoi sen tekeminen tuntua etukäteen jonkin verran epävarmalta toteutuksen kannalta. Haastattelussa tuotiin esiin, että vaikka etukäteen olisi ajatellut toimia tietyllä tavalla, loppujen lopuksi suunnitelmat saattoivat muuttua tutkimuksen edetessä eivätkä suunnitelmat välttämättä muuttuneet käytännöiksi.

Vaikka olisi jokin, kuinka paljon haastatteluja hankitaan ja mistä niitä hankitaan, niin niitä ei välttämättä tule niin paljon kuin haluttiin eikä välttämättä niistä paikoista, mistä suunniteltiin. Että se tietysti jonkin verran vaikuttaa. Ja

sitten on myös sekin jatkokäyttöä ajatellen, että jos sitä haastattelua tulee tietystä organisaatiosta, niin se organisaatio voi sanoa, että näitä ei mihinkään tallenneta ja näitä ei saa jatkokäyttää. Se tulee siinäkin, että ei voi ikään kuin suunnitella, että tämä annetaan jatkokäyttöön. Siinä on tällaisia muuttuvia tekijöitä. (L2)

Suunnittelusta nähtiin kuitenkin myös positiivisia puolia. Suunnitelman tekeminen laittaa edes jonkin verran ajattelemaan tutkimusaineiston käsittelyä jo aivan tutkimuksen alkumetreillä.

Joo, totta kai sillä jokin ryhdistävä vaikutus on, että on pakko ajatella, että mistä tämä aineisto tulee, mitä sille tehdään, miten sitä säilytetään ja mitä sille sitten tapahtuu. Kyllä se jollain tavalla ryhtiä siihen hommaan antaa ja ehkä tosiaan se, että puhutaan nyt ehkä enemmän tästä aineiston hallinnasta ja varsinkin jatkokäytöstä. (L2)

Yksi tutkijoista ei ollut tehnyt aineistohallintasuunnitelmaa koskaan, mutta piti suunnitelmaa silti sellaisena asiana, johon kannattaisi panostaa.

Esimerkiksi tätä viimeistä isoa projektia varten olisi varmaan kannattanut näitä asioita miettiä vielä enemmän, mutta en ole tehnyt mitään sillä lailla muodollista tai oikeastaan epämuodollistakaan. On aina lähtenyt hyvin spontaanisti se aineistonkeruu, ja sitten ne ongelmat tulevat pikkuhiljaa vastaan, mitä siinä sitten tulee. Siinä tutkimusta aloittaessa on harvoin tietoa, paljonko sitä aineistoa tulee sitten olemaan tai mitä saa kasaan. (L4)

Hänen alallaan historian tutkimuksessa tutkijan mukaan aineiston käsittely on jonkin verran omaleimaista verrattuna muuhun tutkimukseen, minkä takia aineistohallinnalle ja sen suunnittelulle tarvitsisi antaa enemmän painoarvoa. Aineistoilla on tutkimusalaan liittyviä erityisongelmia ja -piirteitä myös aineistohallinnan näkökulmasta, ja aineistohallinnan suunnittelu voisi ennalta ehkäistä tutkijoiden yhä uudelleen kohtaamia ongelmia.

Kvantitatiivista tutkimusta tekevä tutkijoista osa oli kuullut aineistohallintasuunnitelman tekemisestä, mutta sellaisen tekeminen ei vaikuttanut olevan yhtä yleistä kuin haastatteleminen kvalitatiivisen tutkimuksen tekijöillä.

Solututkimusta tekevä tutkija kertoi, ettei ole koskaan tehnyt aineistohallintasuunnitelmaa itse tai edes tutkimusryhmän kanssa. Hän kertoi, että tutkimuksessa syntyvät aineistot ovat sen verran pieniä, ettei hallittavuus ole osoittautunut ongelmalliseksi, eikä suunnittelulle ole ollut tarvetta. (M1)

Yksi tutkijoista on ollut vuosien ajan mukana tekemässä seurantatutkimusta, joka on aloitettu 80-luvulla ja ” siinä vaiheessa kukaan ei ollut aineistohallinnasta kuullutkaan”. Tutkimuksen alkuaikoina ei myöskään ollut selvää, että tutkimusta jatkettaisiin, joten aineistohallintaakaan ei tämän aineiston osalta ole juurikaan suunniteltu. (M2)

Eräs tutkija nosti esille myös sen, että tutkimusta aloittaessa on niin paljon erilaisia hallinnollisia papereita täytettävänä ja paljon asioita hoidettavana, että saa tutkimukseen tarvittavat luvat kuntoon. Silloin helposti karsitaan niistä asioista, jotka eivät ole aivan pakollisia, vaikka ne olisivatkin tutkimuksen kannalta hyödyllisiä. Hän kertoi, että jonkinlainen suunnitelma tutkimusaineistojen hallinnasta tutkijoiden on tehtävä, mutta siihen suhtaudutaan helposti vain yhtenä lomakkeena, joka on pakko tehdä. Tutkijan mielestä pitäisi saada kehitettyä sellainen kulttuuri, että tietyt asiat, kuten kunnollinen aineistohallinnan suunnitteleminen, olisivat paljon tärkeämmässä osassa. Hän näkee sen olevan tutkimusta helpottavaa ja erittäin hyödyllistä ja arvokasta tutkimuksen edetessä. (M3)

Osa tutkijoista toi esiin, että aineistohallintasuunnitelman tekoon olisi hyvä saada jonkinlaista apua tai koulutusta. Toive koulutuksesta riippui aika paljon tutkijoiden aineistojen koosta, koska pienten aineistojen hallintaa ja elinkaarta on helpompi suunnitella etukäteen. Osa kertoi, että tällä hetkellä aika pitkälti apua saa kokeneemmilta tutkijoilta ja itse opettelemalla. Muutamat tutkijat kertoivat, että ohjauksen saaminen olisi ollut hyödyllistä esimerkiksi jatko-opintojen aikana, mutta nykytilanteessa eivät osallistuisi koulutukseen, koska se aika on pois muusta tutkimustyöstä. Jotkut tutkijoista kuitenkin kertoivat osallistuneensa aineistohallintaan liittyviin koulutuksiin ja pitäneen niitä hyödyllisinä. (M1, M2) Yksi kertoi törmänneensä monien aloittelevien tutkijoiden kanssa siihen, että tutkimuksen alussa ei ole tarpeeksi tarkasti suunniteltu tutkimuksen kulkua, jolloin myöhemmissä vaiheissa törmää kaikenlaisiin hankaluuksiin niin aineistohallinnan kannalta kuin muidenkin tutkimuksen osa-alueiden kannalta. Hän toivoi, että tutkijat panostaisivat suunnitteluun enemmän ja varsinkin aloittelevat tutkijat pyytäisivät apua jo suunnitteluvaiheessa. (M2)

4.1.2 Kerätyn aineiston käsittely

Käydessämme läpi aineistohallinnan käytäntöjä kerätyn aineiston kanssa, tutkijat toivat esiin esimerkiksi tapoja säilyttää aineistoa tutkimuksen aikana, aineistoon liittyviä rajoituksia, kuten yksityisyydensuoja, sekä muita tapoja toimia aineiston kanssa.

Seuraavassa sitaatissa yksi tutkijoista kuvailee omaa aineistohallintaansa:

Minulle se on se, että olen tehnyt haastattelut ja sitten tallennan ne koneelle salasanan taakse. Äänitykset lähetän litteroitavaksi ja saan tekstitiedostot. Sitten minä printtaan ne itselleni ja säilytän niitä kotona. En halua säilyttää niitä välttämättä yliopistolla, koska haluan lukea niitä kaikessa rauhassa ja silloin luen niitä kotona. Olen luvannut, että säilytän niitä lukitussa kaapissa tai lukitussa tilassa ja minulla ei ole kotona lukittua tilaa ja kaappia, mutta lähden sellaisesta oletuksesta, että kotini on lukittu tila, johon ei tule ihmisiä ja olen ajatellut, että tällainen käytäntö on ihan jees. (...) Ja kun kirjoitan aineistoon perustuen, niin sitten teen sitä jo kirjottaessa, että muutan tunnistettavuustiedot esimerkiksi henkilönimet ja yritän editoida niitä tekstikatkelmia, joissa ne ihmiset ovat yksilöitävissä, ja se on osa sitä työprosessia (L3)

Haastattelemillani laadullisen tutkimuksen edustajilla kaikki aineisto syntyi lähtökohtaisesti sähköisessä muodossa. Osalla oli tapana tulostaa paperille aineistoja, mutta pääosin tutkijat toimivat pelkästään sähköisten aineistojen kanssa. Sähköisiä aineistoja säilytetään pääosin yliopiston verkossa, jossa niiden koetaan pysyvän turvallisesti tallessa. (L1, L2, L3). Yksi tutkijoista toi esiin, ettei pidä pilvipalveluita tarpeeksi luotettavina paikkoina säilyttää tutkimusaineistoja tietoturvan kannalta ja tämä on syy olla käyttämättä pilvipalveluita aineistojen säilytykseen (L2). Vanhoja aineistoja ja vanhoja varmuuskopioita saatetaan säilyttää jonkin verran myös muistitikuilla (L3).

Tutkijat kertoivat jonkin verran myös siitä, minkälaisia tapoja heillä on käsitellä ja jalostaa keräämäänsä aineistoa helpommin käsiteltävään muotoon. Yksi kertoi tekevänsä tutkimusapulaisen kanssa suuresta määrästä haastatteluita eräänlaista Excel-muotoista aineistoarkistoa, johon aineistot on viety tietyllä idealla ja sitä käyttämällä suuresta aineistosta on helpompi löytää tarvittavat tiedot. (L1) Myös toinen tutkija kertoi tekevänsä Excel-muotoista koontia aineistoista. Hänen aineistonsa on pääosin peräisin arkistosta ja arkistosta aineisto on kerätty valokuvaamalla ja siirtämällä kuvat tietokoneelle. Kuvista tut-

kija kokoaa oleelliset sisällöt taulukkoon, johon tekee samalla myös pientä sisällönanalyysiä. Tutkija harmitteli, että hänellä ei riitä aika tekemään koontia niin hyvin ja kattavasti kuin toivoisi. Jos aika riittäisi parempaan ja selkeämpään aineistokoontiin, aineisto olisi sopiva arkistoitavaksi ja siitä voisi olla hyötyä muillekin tutkijoille. (L4)

Kaikki kvantitatiivista tutkimusta tekevät tutkijat kertoivat, että tutkimuksessa syntyy materiaalia sekä sähköisessä muodossa että paperisena. Osalla sitä syntyy vain vähän ja osalla paperiaineistoa tulee paljon kyselylomakkeiden muodossa. Aineistot tallennetaan sähköiseen muotoon manuaalisesti. (M1, M, M3) Paperiaineistoa tulee määrällisesti enimmilläänkin sen verran, että tutkijan on mahdollista tallentaa se itse sähköiseen muotoon, vaikka se onkin aikaa vievää. (M2) Vaikka aineistot tallennetaan sähköiseen muotoon, pääosin kaikki, etenkin kyselylomakkeet, säilytetään varmuuden vuoksi tutkimuksen ajan ja myös tutkimuksen loputtua. Tallennuksessa saattaa tulla virheitä, minkä vuoksi on hyvä säilyttää alkuperäiset aineistot, joista voi tarkastaa syötetyn tiedon oikeellisuutta. (M2, M3) Tutkimuksen aikana papereita säilytetään tutkijoiden omissa huoneissa ja tutkimuksen päätyttyä osa tutkijoista siirtää ne yksikön omiin arkisto- tai varastotiloihin. (M2, M3)

Etenkin sähköisten aineistojen kohdalla on tärkeää pitää huolta siitä, että aineistot pysyvät turvallisesti tallessa. Yhdellä tutkimusryhmällä oli esimerkiksi varmuuskopiointiin yhtenäinen toimintatapa:

Yleinen varmuuskopiointisysteemi, mitä kaikille tiedostoille tehtiin oli, että meille jossain vaiheessa hankittiin jokaiselle omat ulkoiset kovalevyt ja jokaisen omalla vastuulla oli ottaa omista tiedostoistaan säännöllisin väliajoin varmuuskopioita. Sillä tavalla kaikki tutkimusaineisto ja kaikki artikkelit, mitä oli kirjoitettu, eli kaikki tämmöinen data, niin ne niin kuin varmistettiin että ne olivat tallessa (M1)

Aineistoja saatettiin tallentaa myös yhteisesti tutkimusryhmän jäsenten kesken. Yliopiston verkossa on yhteistä levytilaa tutkimusryhmän jäsenille ja jokainen tallentaa aineistonsa sinne kaikkien nähtäville. (M1) Joillain tutkijoilla on tapana myös liikutella aineistoja muistitikuilla ja eräs oli tallentanut vanhoja aineistoja CD-levyille. (M2).

Aineistojen yksityisyydensuojasta huolehtiminen on yksi tärkeä osa aineistohallintaa ja tavat riippuvat tutkittavien kanssa tehdyistä sopimuksista, lainsäädännöstä ja aineiston

arkaluonteisuudesta. Haastattelemillani tutkijoilla oli hyvin erityyppisiä aineistoja, joita koskivat erilaiset rajoitukset.

Eräs tutkija työskentelee pääosin yli sata vuotta vanhojen aineistojen kanssa, joten niihin ei päde enää julkisuulaissa määrätty salassapitovelvollisuus. Toisaalta osa hänen käyttämistään aineistoista on alle sata vuotta vanhoja ja niiden käyttämiseen tarvitsee hakea lupaa. (L4) Toinen tutkija puolestaan kertoi käyttäneensä erääseen tutkimukseen aineistoa, jota ei ollut itse kerännyt vaan saanut sen omaa tutkimusta varten käyttöönsä. Aineiston luovuttaneen osapuolen kanssa tehtiin sopimus siitä, miten aineistoa käytetään ja miten sitä säilytetään turvallisesti. (L2)

Haastattelemillani kvantitatiivisen tutkimuksen tutkijoilla oli kaikilla arkaluonteisia ihmisiin liittyvää terveystietoja, joita koskevat erityiset säädökset. Aineistojen arkaluonteisten tietojen huolelliseen säilytykseen on tarkkoja toimintatapoja. Jotta aineistossa olevat henkilötiedot eivät päätyisi millään tavalla väärin käsiin, on ainoastaan tietyillä henkilöillä oikeus nähdä henkilötietoja, esimerkiksi tutkimusryhmän johtajalla (M1). Muut tutkijat saavat aineiston siten, että tutkimuksen kohteilla on numerokoodit ja aineistosta näkee ne tietyt parametrit, joita tutkijat tarvitsevat laskeakseen tilastollisia yhteyksiä. (M1) Käyttöön liittyvien rajoitusten lisäksi esimerkiksi otosrekisteriä ei saa siirtää yleisen tietoverkon kautta käyttäjältä toiselle. Otosrekisteristä selviää tutkimukseen osallistuneiden henkilötiedot, minkä takia sitä koskevat erityisen tarkat säännöt yksityisyydensuojan turvaamiseksi. (M2)

Aineistoja ei myöskään pääse käsittelemään ihan kuka tahansa tutkija, jos aineisto sisältää potilaisiin liittyvää terveystietoa. Vaikka aineisto olisi anonymisoitu, jokaisella tutkijalla täytyy olla lupa käyttää aineistoa.

Siis tällä hetkellä meidän tutkimusluvut edellyttävät, että vain me käytämme niitä. Jos meille tulee uusi (tutkija), niin heille on pitänyt hakea ihan erikseen oikeudet käyttää, koska meillä on sairaalan eettisen toimikunnan lupa, ja sieltä on myös haettava lupa heille. (M2)

Yksi tutkijoista mietti, että aineiston anonymisoinnista huolimatta voisi periaatteessa olla riskinä se, että joku saattaisi tunnistaa tutkimukseen osallistuneen henkilön. Jos tarpeeksi

monta henkilöön liittyvää tietoa saisi yhdistettyä, saattaisi olla mahdollista tunnistaa tutkittava henkilö, minkä takia aineiston suojeleminen on erittäin tärkeää. (M2)

Muita tapoja suojella tutkittavien henkilötietoja oli esimerkiksi se, että tutkijoiden työhuoneissa ei saa olla papereita, joissa lukee tutkittavien henkilötietoja. Tällaiset paperit hävitetään silppurissa. (M1)

4.1.3 Arkistointi ja jatkokäyttö

Tutkimusaineistojen arkistointi aiheutti jonkin verran keskustelua ja tutkijoilla oli jonkin verran erilaisia näkemyksiä aiheesta. Kukaan haastatteleistani tutkijoista ei ollut hävittänyt tutkimusaineistoja eikä kenelläkään ollut aikomuksena tehdä niin. Aineistojen arkistointia pidettiin kuitenkin ongelmallisena vaikkakin arkistoinnin hyödyt tuotiin myös esiin.

Yleisesti tutkimusaineistojen arkistoinnista kerrottiin, että tutkimuksen loputtua aineistot jäävät usein tutkijoiden omille koneille. Tutkimusaineistot olivat säilyneet tähän asti tutkijoiden koneilla ja vaikka ei ollut suunnitelmaa, mitä niille tulevaisuudessa tapahtuu, hävittämistä ei kuitenkaan nähty vaihtoehtona. (L1, L2, L4, M2, M3) Osa tutkijoista koki huonoa omaatuntoa siitä, että aineistot jäävät vain itselle ja muille tutkimusryhmän jäsenille ja osa mietti, että asialle pitäisi mahdollisesti tehdä jotain jatkossa. Syitä arkistomatta jättämiseen oli esimerkiksi se, että aineistot eivät ole sellaisessa muodossa, että ne voisi toimittaa arkistoitavaksi eikä aineistojen toimittamiselle ollut aikaa riittävästi (L1, L2, L4).

Yksi haastateltavista kertoi olevansa ehdottomasti tutkimusaineistojen arkistoinnin ja jatkokäytön kannalla ja hänellä oli myös selvä suunnitelma olemassa tutkimuksensa aineistojen suhteen.

Niissä on kaikki tunnistetiedot, ja myös siinä vaiheessa, kun olen lahjoittamassa niitä Werstaalle, niin aion lähettää ne sinne sellaisenaan. Osa haastateltavista on toivoneet, että jotain mitä he ovat sanoneet haastattelussa, niin ei tulla käyttämään tutkimuksessa, ja liitän nämä rajaukset siihen lahjoitukseen. Olen kertonut haastateltaville, että en pysty editoimaan sitä äänitiedostoa, että kaikki, mitä ovat sanoneet, niin se tulee äänitiedostona lahjoitetuksi sellaisenaan. Olen myös sanonut, että ne litteroidut tiedostot, että siinä periaatteessa

pystyy tekee editointia, mutta sanon heille, että minulla ei ole siihen resursseja enkä tule tekemään sitä vaan ne lahjoitetaan sellaisenaan ja ne ihmiset, jotka käyttävät aineistoja sitoutuvat käyttämään niitä sellaisella tavalla, että siitä ei tule haittaa näille ihmisille. Tähän mennessä kukaan haastateltavista ei ole kieltäytynyt tällaisesta järjestelystä. (L3)

Tutkimus on historiallista, niin näen sen aineiston anonymisoinnin tapana tuhota tulevaa tutkimusaineistoa ja tehdä siitä tutkimuksellisesti käyttökeltontonta osittain, koska en voi tietää, että miten tutkija, joka käyttää sitä aineistoa, haluaa sitä käyttää. Kuulun siihen koulukuntaan, joka ajattelee, että lahjoitetaan mahdollisimman täydellisenä, ja se on käyttäjän velvollisuus käyttää niitä eettisesti. Ei se ole tutkijan eikä arkistonhoitajienkaan työnkuvan mukaista, sellaista, mistä heille palkkaa maksetaan, että he tuhoavat tutkimusaineistoa. Olen näkemykseltäni ehkä radikaalin avoimuuden kannattaja tässä suhteessa, mutta myös pystyn mielestäni perustelemaan, miksi näin on parempi kuin se että ne anonymisoidaan itse niitä aineistoja. Niiden raporttien anonymisointi on asia erikseen, ja sitä teen ja siihen myös kannustan kaikkia muita, mutta en kannata aineiston anonymisointia. (L3)

Tutkijoilla on tapana säilyttää paperiset aineistot itsellään tutkimuksen aikana ja ne saatavat jäädä tutkijoille myös tutkimuksen päätyttyäkin. Pääsääntöisesti kuitenkin vanhat paperiaineistot menevät säilytykseen erilaisiin arkisto- tai varastohuoneisiin. Yksi tutkijoista mietti, että paperiaineistojen digitointi voisi olla vaihtoehtona, jos aineistoja aiotaan säilyttää pitkiä aikoja, mutta siihen ei ole vielä ryhdytty eikä kunnolla suunniteltukaan (M2). Säilytyksen suhteen ei kenelläkään ollut erityistä suunnitelmaa, mutta paperiaineistojen hävitys ei ainakaan ole ollut aikomuksena. Säilytyksen puolesta puhuu ainakin se, että aineistojen tallennuksessa tapahtuu helposti virheitä, joita täytyy tarkistaa papereista tutkimuksen myöhemmissä vaiheissa. Mahdollisesti niistä täytyy jälkeinpäin käydä hakemassa lisää tietoa, jos alun perin kaikkea tietoa ei ole tallennettu sähköiseen muotoon, jos on ajateltu sen olevan epärelevanttia tutkimuksen kannalta. (M2, M3)

Tässä on juuri käyty suorittamassa arkeologisia tutkimuksia lasten psykiatriaan liittyen, kun on haettu yksityiskohtia, kuten joitakin tiettyjä asioita. Ollaan huomattu, että yhdestä vaiheesta ei oltu tallennettu ainuttakaan päivämäärää, milloin haastattelut on tehty. Piti myös laskea ikää ja todettiin, ettei ole mitään, mistä pisteestä laskea sitä ikää. Sitten huomasimme, että sieltä löytyy informed consent, josta löytyy se tieto, mutta kukaan ei ole ehtinyt penkomaa niitä. Tämän takia ne täytyy olla olemassa, että jossakin vaiheessa keksitään jotain, mitä tarvittiinkin lisää – jotain, mitä ei kukaan ole silloin tullut ajatelleeksi. (M2)

Sähköisiä aineistoja pätee sama periaate, ettei mitään hävitetä. Aineistojen säilyttämisen suhteen ei kuitenkaan ole suunnitelmaa pitkäaikaista säilyttämistä varten, vaan aineistot ovat pääosin tutkijoiden omilla koneilla.

Käytännössä ei ole mitään erityistä arkistointia, niin ne vaan unohtuvat varmaan sinne itse kunkin omalle koneelle. Sitten jos joku tutkija lähtee pois ryhmästä, niin sitten tietenkin pidetään huoli, että kaikki hänen tuottamansa data, mikä on tärkeintä, kopioidaan jonnekin ja jää sille ryhmälle. (M1)

Vanhoja aineistoja on aikoinaan tallennettu myös CD-levyille ja niillä on edelleen jonkin verran aineistoa tallessa. (M3)

Kerättyjä aineistoja saatetaan joissain tapauksissa käyttää myös muuhunkin kuin omaan tutkimukseen. Yksi tutkijoista kertoi, että tutkimuksen aihepiirit voivat olla (lääketieteessä) melko kapeita ja maailmalla kovin moni ei tee samanlaista tutkimusta. Hänen mukaansa tällaisissa tapauksissa on mahdollista, että tutkija pyytää muilta tutkijoilta aineistoja käytettäväksi ja kerää kaiken mahdollisen maailmalla kerätyn tiedon, jotta analysoitava aineisto olisi mahdollisimman suuri ja näin saadaan kerätyt aineistot hyödylliseen käyttöön. (M3)

Vanhat aineistot ovat erittäin arvokkaita myös niissä tilanteissa, jos tulee uusi tutkimusmenetelmä ja aiemmin kerättyä aineistoa voi hyödyntää myös uuden menetelmän kanssa ilman, että ensin tarvitsee kerätä aineistoa uudelleen (M3). Saattaa tulla myös sellaisia tilanteita, että halutaan tutkia aihetta toisesta näkökulmasta, jolloin aikaisemmin tehty tutkimus ja sen aineistot ovat tärkeässä roolissa (M2).

Kvantitatiivisen tutkimuksen edustaja kertoi, että osa tieteellisistä lehdistä vaatii artikke-
lin julkaisun yhteydessä myös aineistoin julkaisua ja nämä ovat ainoita tapauksia, joissa aineistot ovat saatavilla (M3). Haastattelemani kvantitatiivisen tutkimuksen edustajat olivat tekemisissä sen verran arkaluontoisen aineiston kanssa, että aineistojen jatkokäyttö koettiin todella ongelmalliseksi eikä siihen liittyviä toimintatapoja oltu vielä ratkaistu (M2)

4.2 Ongelmista

Tässä luvussa kerron tutkimusaineistojen hallintaan liittyvistä ongelmista, joita tutkijat olivat kokeneet tutkimusaineistojen hallinnan kanssa. Tutkimuksen ollessa käynnissä haasteelliseksi koettiin esimerkiksi aineistojen pitäminen hyvässä järjestyksessä sekä aineistojen arkistointiin ja jatkokäyttöön liittyvät asiat.

4.2.1 Tutkimuksen aikana

Tutkimusaineistojen pitäminen järjestyksessä koettiin joiltain osin ongelmalliseksi. Vaikka kukaan ei kertonut aineistojensa olevan aivan sekaisin tai ettei niistä olisi ollenkaan huolehdittu asianmukaisesti, osa toi kuitenkin esiin sen, ettei ole ihan perillä joidenkin aineistojen sijainnista ja järjestelmällisyydessä nähtiin parantamisen varaa.

Yksi tutkijoista esimerkiksi kertoi, ettei ole varma varmuuskopioidensa sijainnista. Hän kertoi laskeneensa omistavansa ainakin viisi muistitikkoa, joihin on tallennettu tutkimusaineistoja ja varmuuskopioita eri vuosilta ja tutkija arveli, että hänellä itselläänkin voisi olla vaikeaa löytää niiden joukosta tarvitsemaansa, jos se ylipäättään edes löytyisi näiltä tikuilta. Tutkija kuvaili järjestelmällisen aineistohallinnan haasteita seuraavasti:

Laajasti ottaen tutkimuselämään liittyvien data-aineistojen huolellinen hallinta ja tiedostointi tai kansiointi, ja sen järjestelmän laatiminen niin, että sieltä löytyisi joko ajan mukaan tai hankkeiden mukaan tai joidenkin tällaisten mukaan. Vaikka pidän itseäni kohtalaisen järjestelmällisenä ihmisenä, niin minullakin se on ihan retuperällä. Kyllähän usein nämä asiat hoidetaan silloin, kun ne kaatuvat päälle tai tulee joku, että tämä täytyy nyt hoitaa ja sitten se hoidetaan sillä järjellä ja logiikalla, mikä sillä hetkellä tuntuu olevan se päällimmäinen, ja kun tämä kertautee vuosien ja vuosikymmenten ajan, niin kyllähän siitä tulee melkoinen ryteikkö. (L3)

Toinen tutkija puolestaan toi esiin, ettei ole varma, missä tutkimuksen alkuperäiset aineistot ovat:

Ehkä se kertoo projektin huonosta aineistohallinnasta, että minä en tiedä, kenellä on alkuperäiset enkä tiedä, tietääkö sekään, jolla ne alkuperäiset on, että ne ovat hänellä. Se aineisto jotenkin vaan sitten jää, ja on niillä käyttäjillä missä sattuu. (L2)

Järjestelmällisyyden parantamisen puolesta puhui myös kolmaskin tutkija. Hän totesi välillä olevan tilanteita joissa epäjärjestyksessä olevat aineistot hidastavat työtä jonkin verran. Jos aineistot olisivat paremmassa järjestyksessä, jonkin asian voisi tehdä viidessä minuutissa sen sijaan, että pahimmillaan siihen saattaa mennä nykytilanteessa ainakin tunti. (L1)

Aineistojen suuri koko tuntui vaikuttavan siihen, kuinka paljon ongelmia aineistohallinnassa koettiin. Yhdellä tutkijoista tutkimuksessa syntyvät aineistot olivat melko pieniä ja hän ei kokenut aineistohallintaa ongelmalliseksi eikä muistanut törmänneensä aineistohallintaan liittyviin ongelmiin. Ainoastaan se nousi esille, että tutkijat huolehtivat itse aineiston varmuuskopioinnista ulkoisille kovalevyille ja jos sen unohtaa tehdä, se saattaa olla riski. Varmuuskopioiden tekemisessä ei kuitenkaan ongelmia ole tähän asti ollut. (M1)

Sen sijaan varsinkin suurten kvantitatiivisten aineistojen kanssa ongelmaksi saattaa muodostua oman toiminnan puutteellinen dokumentointi, erityisesti aineiston keräämisen vaiheessa. Tutkijat toivat esiin, että suurten aineistojen tai pitkään jatkuvan tutkimuksen aikana on tärkeää dokumentoida kaikki tutkimusta koskevat muutokset ja päätökset, kuten miksi jokin muuttuja on tiputettu pois tai miksi jollekin seurantatutkimukseen osallistuneelle ei ole lähetetty kutsua jatkotutkimukseen. Vaikka tutkijasta hetkellisesti päätökset tuntuvat itsestään selviltä ja tutkimusta aktiivisesti tehdessä tutkimuksen eri yksityiskohdat pysyvät hyvin mielessä, ne saattavat olla unohtuneet ajan kuluessa ja dokumentoimattomasta aineistosta perusteluita muutosten tekemisille ei löydy ja niiden jäljittäminen jälkikäteen on hankalaa. (M2, M3) Yksi tutkijoista kertoi tutkimuspäiväkirjan pitämisen olevan hyvä tapa dokumentoida omaa tekemistään, jotta myöhemminkin löytyy perustelut sille, minkä takia jokin asia on tehty. (M3)

Haastattelemani tutkijat kertoivat, että paperilomakkeiden tiedot tallennetaan manuaalisesti sähköiseen muotoon ja tässä on vaara virheiden tekemiseen. Lomakkeiden kaikkia tietoja ei ole aina tallennettu sähköiseen muotoon sen jälkeen, kun aineisto on kerätty, koska on ajateltu, ettei jotain tiettyä asiaa tarvita tutkimuksessa. Myöhemmin näitä tietoja on saatettukin tarvita, kun on päätetty jatkaa samaa tutkimusta tai aineistoa on käytetty toiseen tutkimukseen. Paperilomakkeet on kuitenkin tähän asti pidetty säilössä, mikä auttaa tilanteissa, jos tietoja ei löydy sähköisessä muodossa. Lomaketietojen manuaalisessa

tallentamisessa on tullut välillä myös virheitä, joiden korjaamiseen tarvitaan paperisia lomakkeita, jossa tieto on. (M2, M3)

Huolellisuuteen ja tarkkuuteen liittyvät asiat nousivat muutenkin esille. Esimerkiksi varmuuskopioita tehdessä saattaa helposti käydä niin, että eri versiot menevät sekaisin ja myöhemmin huomataan, että jokin muuttuja ei olekaan parhaillaan käytettävässä versiossa ja se aiheuttaa hankaluuksia. Yksi tutkijoista oli myös hieman huolissaan aineistojen kuljettamisesta ja siirtelystä muistitikuilla. Vaikka aineistot muistitikuilla ovat anonyymejä, niiden katoamisessa piilee kuitenkin jonkinlainen riski. Lisäksi muistitikkujen säilyminen ehjänä ei ole välttämättä varmaa. Tutkija kertoi kuitenkin, että vaikka erilaisia sekaannuksia on sattunut, niin mitään ei ole koskaan kuitenkaan lopullisesti hävinnyt. (M3)

4.2.2 Tekniset ongelmat

Tutkimusaineistojen hallinnassa nähtiin myös teknisiä ongelmia, jotka koskivat esimerkiksi ohjelmistojä, joita voi käyttää tutkimusaineistojen käsittelyyn. Osa tutkijoista kaipasi avukseen jonkinlaista tutkimusaineistojen hallintaan tarkoitettua ohjelmistoa, jota voisi käyttää apuna tutkimuksessa. Esimerkiksi kuvatiedostojen käsittely koettiin hankalaksi, koska oleellisen tiedon löytäminen kuvatiedostoista käy hitaasti. Tämän tyyppiseen aineiston käsittelyyn ei oikein löydy sopivaa ohjelmistoa ja välillä ohjelmistot myös muuttuvat aiheuttaen hankaluuksia.

En tunne ketään historioitsijaa, jolle olisi selvää esimerkiksi ohjelmisto, miten näitä kuvia sitten käsitellään ja miten sieltä otetaan ne olennaiset, että ne löytyvät nopeasti, vaan kaikki tekevät sen aina joka kerta ja saattavat tehdä eri systeemillä, koska ohjelmistotkin muuttuvat. Se on yksi asia, mikä nyt on ollut tässä viime aikaisessa projektissa ongelmana, että Adoben ohjelma, joka on tavallaan vähän niin kuin kuvagalleriaohjelma, niin sitä ei näissä uusissa Adoben versioissa olekaan. Kotikoneella pystyn käyttäjä sitä ohjelmaa, mutta uudella työkoneellani sitä ei enää ole. (L4)

Tutkijan mukaan aineiston käsittelyä helpottaisi ja nopeuttaisi, jos kuvatiedostoissa olevat tekstit saisi muunnettua tekstitiedostoiksi, mutta tällä hetkellä se ei ole mahdollista kuin pienen määrän kohdalla, koska suuremman aineiston kohdalla se ei ole työekonomisesti mahdollista. (L4)

Aineistojen turvallinen säilyttäminen oli myös yksi aihe, joka herätti keskustelua tutkijoissa. Pilvipalveluiden käyttö aiheutti huolta tutkimusaineistojen turvallista säilytystä ajatellen. Yksi tutkijoista ei halua säilyttää aineistoja yliopiston tarjoamassa pilvipalvelussa, sillä ei koe sen olevan paras paikka arkaluonteista materiaalia ajatellen ja kokee suureksi ongelmaksi ja ristiriidaksi sen, jos kannustetaan pilvipalveluiden käyttöön samalla kun aineistot kuitenkin pitäisi pitää salassa. (L2)

Ei se nyt ole niin, että minulla olisi ihan todella tulenarkaa aineistoa tai materiaalia, mitä pystyisi väärinkäyttämään niin, että se jotain vahingoitettaisi, mutta jos olen luvannut pitää sen aineiston turvassa ja sillä aineistolla tehtävän keskeneräisen tutkimuksen, niin en niitä voi mihinkään pilveen laittaa. (L2)

Teknisenä käytännön ongelmana nousi esiin myös sähköisen aineiston turvallinen siirtäminen. Yksi tutkijoista toi esiin, että verkon kautta ei saa siirtää salassa pidettävää tietoa. Vaikka ihmiset ovat aineistossa koodinumeroilla, tutkija mietti että on hieman siinä rajalla, että ihmisen pystyisi pelkän kerätyn tiedon perusteella tunnistamaan ja silloin aineiston mahdollinen siirtäminen voi olla jonkinlainen ongelma. Todellinen ongelma sen sijaan on otosrekisterin siirtäminen henkilöltä toiselle, sillä otosrekisteristä selviää tutkimukseen osallistuneiden ihmisten henkilöllisyys. (M2)

4.2.3 Arkistoinnin ja jatkokäytön ongelmista

Haastatteluissa kävi selvästi ilmi, että aineistohallinnassa yhdeksi suurimmista ongelmista mainittiin aineistojen arkistointi. Tutkimusaineistot jäävät helposti tutkijoiden omille koneille, eikä niitä välttämättä suunnitelmista huolimatta toimiteta tutkimusaineistoja vastaanottavaan arkistoon. Aineistojen arkistoinnista jättämiseen liittyi monenlaisia syitä ja lisäksi myös tutkimusaineistojen jatkokäyttö nähtiin jonkin verran ongelmalliseksi.

Tutkijat toivat esille, että välillä ei ole aivan selvää, ovatko tutkimusaineistot siinä vaiheessa, ettei niitä käytetä enää. Omia aineistoja halutaan käyttää tutkimuksiin ja artikkeleihin niin kauan, kunnes niistä on saatu aivan kaikki irti, minkä jälkeen ne vasta ovat valmiita arkistoitaviksi. Tähän pisteeseen päätyminen voi olla välillä vaikea huomata, minkä takia aineistot helposti jäävät tietynlaiseen välitilaan, ettei niitä välttämättä käytetä

enää omaan tutkimukseen, muttei kuitenkaan arkistoidakaan ja aineistot jäävät tutkijoiden omille koneille. (L1, L2)

Vaikka yleisellä tasolla tutkijat kokivat tutkimusaineistojen arkistoinnin olevan hyödyllistä ja tärkeää, juuri omat aineistot saatettiin kuitenkin kokea sellaisiksi, joista ei välttämättä olisi hyötyä muille tutkijoille. Yhtenä tekijänä nähtiin esimerkiksi se, että aineistot alkoivat käydä vanhoiksi, jolloin niitä ei voi käyttää tutkimukseen, jonka tarkoitus on kuvata nykyaikaa.

Ne on kerätty vuonna 2008, niin niissä pikkuhiljaa voi tulla sellainen, että ne ovat jo vähän vanhoja. Eli se niiden arkistointi (mietityttää), että hyötyykö niistä kukaan sitten tässä vaiheessa, kun ne on pikkaisen liian vanhoja siihen, että niistä voisi sanoa, että nykypäivän työelämä on näin, vaikka ei se nyt niin nopeasti muutu. Mutta ei välttämättä joku graduntekijä halua tarttua sellaiseen aineistoon, joka on vähän siinä rajalla, että joko se haiskahtaa, mutta ei ole tarpeeksi vanhaa, että voisi sanoa, että ennen oli näin ja nyt on näin. (L2)

Toisaalta tutkija toi myös esille sen, että myöhemmin tulevaisuudessa vertailevaa tutkimusta voisi tehdä, jos tutkijat arkistoisivat aineistojaan, vaikka tuntuu ettei niistä juuri tällä hetkellä olisi kenellekään hyötyä. Hän kertoi, että esimerkiksi nyt olisi todella mielenkiintoista saada käsiin 70-luvun aineistoja. (L2)

Ongelmia nähtiin myös aineiston muodossa. Historiaa tutkiva kertoi, että monet hänen aineistoistaan, lukuun ottamatta haastatteluaineistoja, löytyvät jo valmiiksi arkistoista, minkä takia hänellä ei ole samanlaista painetta toimittaa omia aineistojaan arkistoon kuin jos olisi kyseessä sellaiset aineistot, joita ei mistään vielä löydy. Etuna tutkijan omissa aineistoissa on se, että niissä on yhteen kerättynä tietystä aiheesta tietoa, kun nykyisellään se on ripoteltuna ympäri arkistoa/arkistoja ja aihetta mahdollisesti seuraavaksi tutkiva joutuu tekemään samanlaisen työn etsiessään ja yhdistäessään tietoja. Tutkijan oman aineistot ovat kuitenkin hyvin pitkälti suuri määrä valokuvatiedostoja arkistoaineistoista sekä Excel-taulukkoja ja tutkijan mukaan tässä tilassa olevasta aineistosta saatava hyöty on kyseenalainen, ja jotta siitä olisi hyötyä muille, se pitäisi saattaa parempaan muotoon ja järjestykseen ennen arkistoon toimittamista. (L3)

Toinenkin tutkija mainitsi, ettei tutkimusaineistot ole tällä hetkellä sellaisessa järjestyksessä ja muodossa, että ne voisi suoraan toimittaa arkistoitavaksi. Tutkimusaineistojen

laittaminen sellaiseen kuntoon, että ne voitaisiin arkistoida, koettiin aikaa vieväksi tehtäväksi:

... siihen menee paljon aikaa ja vaivannäköä ja näin ja sitten pitäisi olla vaikka joku erillinen rahoitus siihen osoitettuna tai et se voisi olla jonkun väikkärintekijän, joka on siinä projektissa mukana, niin sen hommiin kuulua se. Mutta tavallaan sitä ottaa omasta työajastaan siihen arkistointiin erikseen vielä aikaa, niin se on vähän haastavaa. (L1)

Yksi tutkijoista toisaalta ei kokenut tutkimusaineistojen arkistointia sillä tavalla hyödylliseksi, että niitä voisi sellaisenaan käyttää uuteen tutkimukseen, ellei alusta asti ole suunnitellut tutkimusta sen pohjalta, että tietää aiheesta olevan tutkimusaineisto tallennettuna.

Tutkimusaineistojen arkaluonteisuus on yksi isoimmista ongelmista tutkimusaineistojen arkistoinnin ja jatkokäytön kannalta. Anonymisointi ei välttämättä riitä poistamaan aineistosta tutkittavien tunnistettavuutta. Erään tutkijan mielestä on hankala arvioida, kuinka paljon aineistosta pitäisi poistaa (tunniste)tietoja, jotta tutkittavien anonymiteetti varmasti säilyisi. Poistamiset puolestaan vaikuttaisivat aineiston laatuun negatiivisesti. Kaikkein hankalimpia aineistoja ovat videoaineistot, koska niistä ei pysty eliminoimaan tutkittavaa pois millään tavalla ja vielä ei ole keksitty ratkaisua sille, miten videoaineistojen kanssa voisi parhaiten toimia arkistoinnin kannalta. (M2)

Yksi tutkijoista oli ratkaissut anonymiteettiin liittyvän ongelman sillä, että arkistoi aineistonsa sellaiseen paikkaan, jossa aineistojen käyttö ei ole kaikille vapaata, vaan niihin annetaan käyttöoikeuksia tietyin perustein. Tällöin aineiston voi luovuttaa ilman anonymisointia. Tutkijat kokivat, että joissain tapauksissa aineiston laatu heikkenee huomattavasti, jos siitä poistaa kaikki tunnistetiedot.

Minulla on sellainen pulma, että aineistoissani on nyt tilallisuus aika paljon läsnä, kontekstuaalisuus. Ja sitten tutkin kokemuksellista maailmaa. Että tavallaan kokemuksellisuus tuottaa siihen hirveän henkilökohtaisen ulottuvuuden niihin aineistoihin, ja maantieteellisyys tuottaa sen, että ne ovat hyvin helposti paikannettavissa. Niin sitten jos miettis, että lähtisin keräämään aineistoa nyt joidenkin no vaikka lasten tai nuorten kanssa tai vaikka – mitä me ollaan kehittelemässä yhdessä ryhmässä maahanmuuttajatutkimusta vaikka turvapaikanhakijoiden kanssa – niin miten minun pitäisi suunnitella se, jotta sen voisi anonymisoida ja että siinä edelleen olisi jotain järkeä. Niin aika haastavaa se olisi, koska niissä tultaisiin puhuman sen ihmisen kokemusmaailmasta, ja tyypillistä on tämäntyyppiset keskustelut: no, silloin ja silloin, kun minä asuin tuolla ja tuolla, niin tässä lähellä oli ja kävin usein siellä ja siellä

ja se oli sillä tavalla hankalaa, kun vanhempani asuivat siellä ja siellä ja välimatka oli sen verran ja niin. On tavallaan aika vaikeaa ottaa pois se konteksti ilman, että siitä häviää todella paljon. Ja sitten toisaalta ne ovat monesti niin henkilökohtaisia, että jos joku sanoo, että mikä asuinalue ja mitä harrastaa ja vaikka että, missä maassa on sukulaisia, niin ihminen alkaa olla aika pieneen ruutuun laitettu. Sekin on yks syy siihen, että tuon tyyppinen aineisto vaatisi aika paljon työtä ja miettimistä aika paljon etukäteen (L1)

Sen sijaan historiallisia aineistoja ei koettu yksityisyyden suojan kannalta ongelmallisiksi, koska haastatteleman tutkija työskenteli pääasiassa yli sata vuotta vanhojen aineistojen kanssa, joita ei koske enää yksityisyyden suoja. (L3) Myös jonkin alan asiantuntijoiden haastattelut koettiin sellaisiksi, että ne on helppo toimittaa arkistoitaviksi ja niitä voi huoletta käyttää, sillä asiantuntijat usein antavat luvan arkistointiin ja haastattelut eivät usein sisällä niin arkaluonteista materiaalia kuin tavalliset henkilöhaastattelut, sillä asiantuntijat puhuvat heidän työstään ja lähtökohtaisesti sellaisia asioita, joita voisi omalla nimellä sanoa. (L1)

Kvantitatiivisen tutkimuksen edustajat kertoivat, että sähköisiä aineistoja ei ole tällä hetkellä toimitettu minnekään tutkimusaineistoja vastaanottavaan arkistoon. Yksi tutkijoista mietti, että jonkinlainen arkistointisuunnitelma pitää jossain vaiheessa kehittää viimeistään siinä vaiheessa, kun hän on jäämässä eläkkeelle tai muista syistä lopettamassa tutkimuksen parissa. Hän mietti, että tällä hetkellä ei taida oikein löytyä oikeanlaista ratkaisua todella arkaluonteisten tutkimusaineistojen arkistoinnille, mutta toivoi arkistointisysteemien kehittyvän vielä tulevaisuudessa ja ratkaisun löytyvän. Tutkija mietti, että paras tapa olisi sellainen arkisto, jonne aineistot voisi luovuttaa anonymisoimatta ja aineiston käyttöoikeuksien kanssa oltaisiin todella tarkkoja. Esimerkiksi sillä tyylillä, että arkistolla olisi tiedot tutkittavista henkilöistä, mutta käyttöluvan saaneille tutkijoille aineisto luovutettaisiin käyttöön anonymisoituna (eli tutkittavat henkilöt olisi numerokoodattu). Käyttö lupaan pitäisi olla kuitenkin perusteltu syy, sillä anonymisoidun aineistonkin kanssa voisi olla pieni mahdollisuus siihen, että kaikki tiedot yhdistämällä tutkittavan henkilön voisi tunnistaa. Tutkijan mielestä kuitenkin tarkkaan mietitty arkistointisysteemi pitäisi aineistot paremmassa turvassa kuin nykyisen kaltainen tilanne, jossa ne ovat tutkijoiden omilla koneilla. (M2)

Yksi tutkijoista koki haitalliseksi ainakin omien aineistojensa jatkokäytössä sen, että niistä voi helposti tehdä vääränlaisia tulkintoja. Hän pohti sitä, miten aineistot voisi saada

sellaiseen muotoon, ettei virhetulkintoja tapahtuisi ja että mahdollinen aineistojen seuraava käyttäjä ymmärtäisi aineiston taustalla olevat tekijät.

Ei pitäisi käydä mitään sellaista, että sitten myöhemmin tulee miettineeksi, että joku muu tulkitsee tätä ja se voi saada jonkun ihan väärän käsityksen ja tehdä jotain ihan hulluja tulkintoja siitä, minkä joku on kanssani jakanut jonkun henkilökohtaisen ajatuksen ja niin, että minä olen keskustellut siitä hänen kanssaan tai tunnen häntä muuten niin, että voin ymmärtää mitä hän sillä tarkoittaa. (L1)

Tutkija oli huolissaan siitä, että aineistoa käytettäisiin seuraavan tutkijan jonkin argumentin perustelemiseen, vaikka aineiston perusteella tehty tulkinta olisi väärä. Pelkän aineiston perusteella ei välttämättä saa tietoonsa sellaisia aineiston tulkintaan vaikuttavia tekijöitä, kuin mitä tutkijalla on sen perusteella, kun on tavannut henkilökohtaisesti haastateltavan.

Toinenkin tutkija toi esiin, että pelkästään tutkittavien anonymiteetti ei ole arkistoinnin ja jatkokäytön ongelma. Tutkijan mielestä mahdollisen seuraavan käyttäjän pitäisi ymmärtää hyvin käyttämäänsä aineistoa, esimerkiksi käytettyjä tutkimuksessa käytettyjä menetelmiä, jotta pystyy kunnolla käyttämään aineistoa. Tämän lisäksi alkuperäisen tutkijan olisi pitänyt dokumentoida aineisto niin hyvin, että seuraava tutkija, joka ei tiedä aineistosta mitään etukäteen, voisi ymmärtää aineiston sisältöä. (M2)

Osa kvalitatiivisen tutkimuksen edustajista nosti esiin myös eräänlaisen mustasukkaisuuden tutkimusaineistoja kohtaan. Tällaista oltiin havaittu sekä omalla kohdalla, että yleisesti tutkijoiden keskuudessa. Kun aineiston keräämiseen on nähty paljon vaivaa, niitä ei välttämättä halutakaan antaa muiden käyttäväksi.

Ehkä sellainen jonkinlainen tutkijan mustasukkaisuus voi olla, että nämä ovat suurella vaivalla kerättyjä ja me tehtiin tässä paljon töitä, ja nytkö nämä pitäisi vaan antaa tänne, että ehkä sitten senkin takia voi olla, että hautoo niitä tai miettii, että mikä se on se oikea vaihe antaa ne pois. (L2)

Ainakin sitten halutaan tehdä niin, että on itse julkaissut siitä kaiken, mitä haluaa julkaista sen sijaan, että siinä kohtaa, kun itsekin vielä tietää, että aikoo vielä monta vuotta työskennellä saman aiheen parissa, niin antaisi muille mahdollisuuden käyttää sitä samaa aineistoa, niin voi tulla se olo, että ettei tämä ole mitään hyväntekeväisyyttä vaan että tämä on ihan tällaista työtä. (L1)

Kyllä luulen, että joillakin ihmisillä on aika vahvasti sellaista, että kun minä näen hirveän ison vaivan, ja olen saanut hyvän rahoituksen ja kerännyt omasta näkökulmasta todella hyvän aineiston, ja sitten että sen antaisi ilmaiseksi muille, niin siinä voi olla henkistä hankaluutta. (L1)

Yksi tutkijoista kertoi myös, että omaa aineistoa saattaa hieman hävetä, minkä takia sen luovuttaminen arkistoon tuntuu vaikealta. Hän sanoi, että helposti mieltii, onko haastatteluissa jotain selkeitä epäonnistumisia, tai ajatteleeko aineiston lukija, että tutkija ei ole saanut haastateltavasti mitään irti ja onko käynyt niin, että haastatteliija puhuu haastateltavan päälle koko ajan ja keskeyttelee. Toisaalta hän kyllä tiedosti sen, ettei aineistoihin jää välttämättä edes näkyviin kuka on haastatellut ja vaikka näkyisikin, niin olisiko sillä loppujen lopuksi kovin suurta merkitystä. (L2)

Tutkijan ”mustasukkaisuus” tuli esille myös kvantitatiivista tutkimusta tekevien haastatteluissa. Eräs tutkija tulkitsi tilannetta niin, että kun tutkimusta tehdessä on nähnyt suuren vaivan rahoituksen saamiseen ja aineiston keräämiseen, saattaa helposti tulla sellainen olo, että haluaa ottaa ensin aineistostaan maksimaalisen hyödyn irti, ennen kuin antaa muiden hyödyntää keräämäänsä aineistoa. Yhtenä tekijänä saattoi tutkijan mielestä olla tämän maan rahoituspolitiikka ja tutkimuksen eteen nähty massiivinen työmäärä, minkä takia omista tekemisistään halutaan pitää tiukasti kiinni. (M3)

5 POHDINTAA JA LOPPUPÄÄTELMÄT

Tutkimukseni tavoitteen oli selvittää tutkijoiden aineistohallintatapoja sekä tuoda esiin mahdollisia aineistohallintaan liittyviä ongelmia. Tässä luvussa teen yhteenvedon tutkimustuloksistani.

Aineistohallinnasta puhuttaessa usealle tutkijalle tuli mieleen pääosin aineistojen arkistointiin ja jatkokäyttöön liittyvät asiat. Haastatteluissa tavoittelin kuitenkin laajemmin tutkimusaineiston elinkaareen läpikäyntiä aineistohallinnan suunnittelusta aineiston jatkokäyttöön.

Kuten kirjallisuuskatsauksessa kerroin, aineistohallinta lähtee jo tutkimuksen suunnitteluvaiheesta, jolloin on mahdollista suunnitella myös tutkimukseen liittyvän aineiston elinkaarta. Suunnitteluvaiheessa tehdyt ratkaisut kulkevat mukana läpi tutkimuksen, joten aineistohallinnan kannalta tässä vaiheessa on tärkeää miettiä aineiston käyttöön, säilyttämiseen ja mahdolliseen jatkokäyttöön liittyvät toimenpiteet.

Osalla tutkijoista oli kokemusta aineistohallintasuunnitelman tekemisestä. Suunnitelman teon merkitys korostui etenkin suurten aineistojen kanssa toimivilta, jolloin mahdollisten huonojen menetelmien aiheuttamien ongelmien korjaaminen on työläämpää kuin pienempien aineistojen kanssa toimivilla.

Toisaalta aineistohallinnan suunnittelu ei poista kaikkia aineistoon liittyviä ongelmia. Vaikka suunnitelman teolla koettiin olevan jonkinlainen ryhdistävä tekijä aineistohallinnan suunnittelun kannalta, kaikille ei ollut aivan selvää minkälaisia asioita suunnitelmaan tulisi sisällyttää, jolloin suunnitelmat tehtiin melko pintapuolisesti. Steinhartin (2012) tutkimuksessakin tuli esiin, etteivät tutkijat välttämättä tiedä, minkälaisia asioita aineistohallintasuunnitelmassa tulisi käsitellä ja tutkijat toivoivat myös koulutusta suunnitelmien tekoon. Hankaluutta aiheutti esimerkiksi se, etteivät tutkijat osanneet nimetä aineistohallinnan eri osa-alueita ja toisaalta oli myös hankala tunnistaa mikä syntyvästä aineistosta on itse tutkimusainestoa ja mikä taas jonkinlaista oheismateriaalia. (Steinhart ym. 2012) Omassa tutkimuksessani suunnitelman tekemisen koki helpoimmaksi tutkija, jolle aineistohallinnan osa-alueet olivat jossain määrin tuttuja ja jolle esimerkiksi tutkimusaineiston tallentaminen arkistoon ja sen jatkokäytön mahdollistaminen oli tärkeää.

Tällöin suunnitelmaan on helppo avata, miten aineistoa käsittelee ja miten sen tallentaa. Toisaalta nousi myös esiin, että vaikka suunnitelmaan oli kirjoittanut, että aikoo toimittaa aineistonsa arkistoitavaksi tutkimuksen päätyttyä, se ei kuitenkaan ole ohjannut tutkijan toimintaa aineistojen kanssa, sillä aineistot ovat jääneet tutkijan koneelle ja ovat sellaisessa muodossa ja järjestyksessä, että vaatisi suuren työn muuttaa ne arkistointikelpoiksi.

Tutkijoilla syntyi aineistoa pääosin sähköisesti, mutta osalla tutkijoista aineistoa syntyi myös paperisena, jotka keräämisen jälkeen tallennettiin manuaalisesti sähköiseksi. Aineistoa säilytettiin pääosin yliopiston palvelimilla, mutta jonkin verran tunnistetietoja sisältämättömiä aineistoja säilytettiin tai kuljetettiin muistitikuilla ja joitakin vanhoja aineistoja oli tallessa CD-levyillä.

Aineiston turvallisen säilyttämisen periaatteet tutkimuksen aikana olivat tutkijoille tuttuja. Hyvä tieteellinen käytäntö ohjaa, että tutkimuksessa syntyneet aineistot tallennetaan tieteelliselle tiedolle asetettujen vaatimusten mukaisesti (Hyvä tieteellinen käytäntö). Tutkimuseettisesti tärkeä periaate on yksityisyyden suoja ja tutkimusaineistojen kohdalla tärkein yksityisyyden suojan osa-alue on tietosuojat. (Ettinen ennakkoarviointi ihmistieteissä) Henkilötietolaissa säädetään tunnistellisen tiedon käsittelystä (523/1999) ja perusperiaatteita ovat, että henkilötietojen käsittely on sallittua tutkittavan luvalla ja tunnistellisia aineistoja voi kerätä ja käyttää silloin, kun se on tutkimuksen kannalta tarkoituksen mukaista. Lisäksi huolehditaan, että tunnistellista aineistoa säilytetään luottamuksellisesti ja suojattuna. (Ettinen ennakkoarviointi ihmistieteissä)

Juuri yksityisyyden suojaan liittyvät asiat aiheuttivat pohdintaa tutkijoissa. Tämä nousi esiin niin aineiston käsittelyyn kuin jatkokäyttöön liittyvän puheen yhteydessä. Tutkijat pohtivat esimerkiksi voiko pilvipalveluissa säilyttää arkaluonteista tutkimusmateriaalia ja miten aineistoa siirretään turvallisesti tutkijalta toiselle. Arkistointiin ja jatkokäyttöön liittyvät pohdinnat puolestaan liittyivät siihen, voiko aineistoa ylipäättään käyttää kukaan muu kuin alkuperäinen tutkija aineiston arkaluonteisuuden vuoksi tai jos aineiston anonymisointi pilaa aineiston laadun.

Lähes kaikilla haastattelemillani tutkijoilla aineistot jäivät omille tai tutkimusryhmän eri jäsenten tietokoneille tutkimuksen loputtua. Higgins (2012) toi esiin, että tutkimusaineistoja ei välttämättä tarvitse tallentaa varsinaiseen arkistoon, vaan säilytyspaikaksi käy myös esimerkiksi yliopiston palvelin, kunhan aineistoa hallitaan siten, että ne ovat käytökelpoisia pidemmänkin ajan jälkeen ja niihin on pääsy myös muillakin kuin alkuperäisellä tutkijalla (Higgins 2012).

Tutkimuksessani ilmeni samanlainen ongelma laadullisten aineistojen jatkokäyttöön liittyen kuin Parryn ja Mauthnerin tutkimuksessa (2004). Tietynlaisissa laadullisissa aineistoissa tunnistetietojen hävittäminen vaatii aineiston suurta karsimista, joka huonontaa aineiston laatua huomattavasti jatkokäyttöä ajatellen. Yksi haastateltavistani oli ratkaissut ongelman siten, että tallentaa tunnistetietoja sisältävän aineiston arkistoon ja aineiston jatkokäyttöä rajoitetaan käyttöoikeuksin. Tällöin aineiston laatu ei kärsi ja mikäli tutkitavat henkilöt antavat tähän suostumuksensa, myös tämä ratkaisu on mahdollinen (henkötietolaki).

Borgin ja Kuulan (2007) tutkimuksessa tietosuojaan liittyvät asiat olivat yksi ongelma-alueista ja samassa tutkimuksessa nousi esiin myös muitakin samoja teemoja, joita tutkijat toivat esiin omassa tutkimuksessani, kuten tutkijan mustasukkaisuus omasta aineistosta, puutteellinen ohjeistus yliopiston taholta ja se, ettei omaa tutkimusaineistoa välttämättä osaa tulkita oikein kukaan muu kuin alkuperäinen tutkija.

Kaiken kaikkiaan aineistohallinta on monimuotoinen aihealue, johon liittyy useita osa-alueita ja näkökulmia. Tutkijoille oli osittain vaikea tunnistaa, mitä kaikkea aineistohallintaan kuuluu, mutta toisaalta aineistohallinta oli heille joiltain osin kuitenkin tuttua. Aineistohallintaan on tarjolla enenevässä määrin koulutusta ja se on paljon esillä esimerkiksi Yhteiskuntatieteellisen tietoarkiston seminaareissa ja koulutuksissa, joten se tekee aiheesta tutummaksi tutkijoille tulevaisuudessa. Tutkimuksessa esiin nousseet aineistohallintaan liittyvät ongelmat olivat pitkälti saman tyyppisiä kuin aiemmin tehdyssä tutkimuksessa ja aiheesta riittää vielä tutkittavaa ajan kuluessa, mikäli aineistohallintaan liittyvät käytännöt yleistyvät tai muuttuvat.

Jälkikäteen tutkimustani arvioidessani mietin, että haastateltavien valinnalla on ollut merkitystä tutkimuksessa esiin nouseviin asioihin. Tutkimustulos olisi voinut olla erilainen,

mikäli olisin haastatellut enemmän tutkijoita tai vain sillä, että olisin valinnut toiset henkilöt haastateltaviksi. Haastattelin sekä kvalitatiivista että kvantitatiivista tutkimusta tekeviä tutkijoita. Alun perin olin ajatellut verrata näitä ryhmiä keskenään, mutta haastateltavien pienen lukumäärän vuoksi koin, etteivät tulokset olleet vertailukelpoisia, etenkin kun kvantitatiivista tekevät tutkijat tekivät melko saman tyyppisiä tutkimuksia ja aineistohallintaa ohjasivat paljon myös tiettyyn alaan liittyvät säädökset, jotka eivät ole yleistettävissä.

LÄHTEET

Borg, S., ja Kuula, A. 2007. Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari.

Cox, A. M., ja Pinfield, S. 2014. Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316. <http://doi.org/10.1177/0961000613492542>

CSC– Tieteen tietotekniikan keskus Oy. 2010. Tutkimuksen tietoaaineistot. Olennaisen käsikirja päättäjille.

Dietrich, D., Adamus, T., Miner, A., ja Steinhart, G. 2012. De-mystifying the data management requirements of research funders. *Issues in Science and Technology Librarianship*, 70. <http://doi.org/10.5062/F44M92G2>

Doorn, P., ja Tjalsma, H. 2007. Introduction: Archiving research data. *Archival Science*, 7, 1–20. <http://doi.org/10.1007/s10502-007-9054-6>

Eettinen ennakkoarviointi ihmistieteissä. Tutkimuseettinen neuvottelukunta TENK. <http://www.tenk.fi/fi/eettinen-ennakkoarviointi-ihmistieteissa>

Eskola, J., ja Suoranta, J. 1998. Johdatus laadulliseen tutkimukseen. Tampere: Vastapaino.

Gilliland, A. 2014. Conceptualizing 21st-century archives. Chicago: Society of American Archivists.

Henkilötietolaki 523/1999. <http://www.finlex.fi/fi/laki/ajantasa/1999/19990523>

Henttonen, P. 2015. Johdatus asiakirjahallinnan tutkimukseen. Helsinki: Avain.

Higgins, S. 2012. The lifecycle of data management. In *Managing Research Data*. London: Facet.

Hirsjärvi, S., ja Hurme, H. 2008. Tutkimushaastattelu: teemahaastattelun teoria ja käytäntö. Helsinki: Gaudeamus Helsinki University Press.

Hyvä tieteellinen käytäntö. Tutkimuseettinen neuvottelukunta TENK.

<http://www.tenk.fi/fi/hyva-tieteellinen-kaytanto>

Hirsjärvi, S., Remes, P., ja Sajavaara, P. 2009. Tutki ja kirjoita. Helsinki: Tammi.

Laine, J. 2009. Tiedonhallintasuunnitelma. Mitä, miksi ja miten?, 1–6.

Parry, O., ja Mauthner, N. S. 2004. Whose Data Are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology: The Journal of the British Sociological Association*, 38(1), 139–152.

<http://doi.org/10.1177/0038038504039366>

Pennock, M. 2007. Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information. *Library & Archives*, (1).

Pinfield, S., Cox, A. M., ja Smith, J. 2014. Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLOS ONE*, 9(12), e114734.

<http://dx.doi.org/10.1371%2Fjournal.pone.0114734>

Pryor, G. 2012. Managing research data. London: Facet.

Ray, J. M. 2014. Research Data Management : Practical Strategies for Information Professionals. West Lafayette, Indiana: Purdue University Press.

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., ja Kramer, S. 2012. Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements. *Journal of eScience Librarianship*.

Suomen Akatemia. 2015. Avoin tiede. <http://www.aka.fi/fi/rahoitus/hyva-tutkimus-tapa/tieteen-avoimuus/>

Sykes, J. 2008. Managing the UK's Research Data: Towards a UK Research Data Service. *New Review of Information Networking*, 14, 21–36.

<http://doi.org/10.1080/13614570902953556>

Tampereen yliopiston arkistonmuodostussuunnitelma. Viitattu 18.3.2018. Tampereen yliopisto. Saatavilla Tampereen yliopiston intrasta.

Tuomi, J., ja Sarajärvi, A. 2009. Laadullinen tutkimus ja sisällönanalyysi.

Vines, T., Andrew, R., Bock, D., Franklin, M., Gilbert, K., Kane, N., ... Yeaman, S. 2013. Mandated data archiving greatly improves access to research data. FASEB Journal, 27, doi: 10.1096/fj.12-218164 fj.12-218164. <http://doi.org/10.5061/dryad.6bs31>

Yakel, E. 2007. Digital curation. OCLC Systems & Services. <http://doi.org/10.1108/10650750710831466>

Yhteiskuntatieteellinen tietoarkisto. Aineistohallinnan käsikirja. <http://www.fsd.uta.fi/aineistohallinta/fi/>